

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.				
1. AGENCY USE ONLY (Leave blank)		2. REPORT DATE 14.Aug.02	3. REPORT TYPE AND DATES COVERED DISSERTATION	
4. TITLE AND SUBTITLE WAVELET DOMAIN IMAGE RESTORATION AND SUPER-RESOLUTION			5. FUNDING NUMBERS	
6. AUTHOR(S) MAJ GODA MATTHEW E				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) UNIVERSITY OF ARIZONA			8. PERFORMING ORGANIZATION REPORT NUMBER CI02-119	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) THE DEPARTMENT OF THE AIR FORCE AFIT/CIA, BLDG 125 2950 P STREET WPAFB OH 45433			10. SPONSORING/MONITORING AGENCY REPORT NUMBER	
11. SUPPLEMENTARY NOTES				
12a. DISTRIBUTION AVAILABILITY STATEMENT Unlimited distribution In Accordance With AFI 35-205/AFIT Sup 1			12b. DISTRIBUTION CODE	
13. ABSTRACT (Maximum 200 words)				
<div style="display: flex; justify-content: space-between; align-items: center;"> <div style="text-align: center;"> DISTRIBUTION STATEMENT A Approved for Public Release Distribution Unlimited </div> <div style="font-size: 2em; font-weight: bold;">20020829 048</div> </div>				
14. SUBJECT TERMS			15. NUMBER OF PAGES 148	
			16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT	

WAVELET DOMAIN IMAGE RESTORATION AND SUPER-RESOLUTION

Matthew Elmer Goda, Ph.D.
The University of Arizona, 2002

Director: Bobby R. Hunt, Ph.D.

Multi-resolution techniques, and especially the wavelet transform provide unique benefits in image representation and processing not otherwise possible. While wavelet applications in image compression and denoising have become extremely prevalent, their use in image restoration and super-resolution has not been exploited to the same degree. One issue is the extension 1-D wavelet transforms into 2-D via separable transforms versus the non-separability of typical circular aperture imaging systems. This mismatch leads to performance degradations.

Image restoration, the inverse problem to image formation is the first major focus of this research. A new multi-resolution transform is presented to improve performance. The transform is called a Radially Symmetric Discrete Wavelet-like Transform (RS-DWT) and is designed based on the non-separable blurring of the typical incoherent circular aperture imaging system. The results using this transform show marked improvement compared to other restoration algorithms both in Mean Square Error and visual appearance. Extensions to the general algorithm that further improve results are discussed.

The ability to super-resolve imagery using wavelet-domain techniques is the second major focus of this research. Super-resolution, the ability to reconstruct object information lost in the imaging process, has been an active research area for many years. Multiple experiments are presented which demonstrate the possibilities and problems associated with super-resolution in the wavelet-domain. Finally, super-resolution in the wavelet domain using Non-Linear Interpolative Vector Quantization is studied and the results of the algorithm are presented and discussed.

WAVELET DOMAIN IMAGE RESTORATION AND
SUPER-RESOLUTION

by
Matthew Elmer Goda

A Dissertation Submitted to the Faculty of the
DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING
In Partial Fulfillment of the Requirements
For the Degree of
DOCTOR OF PHILOSOPHY
In the Graduate College
THE UNIVERSITY OF ARIZONA

2002

THE UNIVERSITY OF ARIZONA ®
GRADUATE COLLEGE

As members of the Final Examination Committee, we certify that we have
read the dissertation prepared by Matthew Elmer Goda
entitled Wavelet Domain Image Restoration And Super-Resolution

and recommend that it be accepted as fulfilling the dissertation
requirement for the Degree of Doctor of Philosophy

B. R. Hunt
Bobby R. Hunt, Ph.D.

6/19/02
Date

Robert A. Schowengerdt
Robert A. Schowengerdt, Ph.D.

6/19/02
Date

Michael M. Marcellin
Michael M. Marcellin, Ph.D.

6/19/02
Date

William J. Dallas
William J. Dallas, Ph.D.

6/19/02
Date

Date

Final approval and acceptance of this dissertation is contingent upon
the candidate's submission of the final copy of the dissertation to the
Graduate College.

I hereby certify that I have read this dissertation prepared under my
direction and recommend that it be accepted as fulfilling the dissertation
requirement.

Bobby R. Hunt
Dissertation Director Bobby R. Hunt, Ph.D.

6/19/02
Date

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at The University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

The views expressed in this article are those of the author and do not reflect the official policy or position of the United States Air Force, Department of Defense, or the U.S. Government.

SIGNED: _____

A handwritten signature in dark ink, appearing to read "W. M. J. Goda", is written over a horizontal line.

ACKNOWLEDGEMENTS

I'd like to express my appreciation to some of the many people who have helped me during the past three years. First, to Dr. Bobby Hunt for providing a great environment to work in. You have been a pleasure to work under, learn from and interact with. Also, Dr. Michael Marcellin for providing the best classes I had the privilege to attend here at the University as well as many insightful discussions along the way. Many students have provided help and support along the way including Joe Green who was always willing to go out for lunch and talk things over; Jim Parco who filled in later on; Bryn Davis who led me to appreciate New Zealand; and Ali Bilgin for discussions on VQ and related topics.

An especially deep note of appreciation is due to my wife for the love and encouragement she provided every step of the way during this program, and also for the past ten years that brought us to this point. Although she may not understand much past this point, she fills my life with joy in a way nothing else can.

Finally I want to thank the Creator of all we see and that which I study, and his son, Jesus, who sustains me daily. 'It is the glory of God to conceal a matter; to search out a matter is the glory of kings.' Proverbs 25:2 (NIV) It is glorious to ponder the way He has set things in motion and continues to sustain them!

TABLE OF CONTENTS

LIST OF FIGURES	7
LIST OF TABLES	12
ABSTRACT	13
CHAPTER 1. INTRODUCTION	14
1.1. Problem Definition	14
1.2. Research Goals	14
1.3. Outline	15
CHAPTER 2. FUNDAMENTALS OF IMAGE FORMATION AND RESTORATION	17
2.1. Overview	17
2.2. Image Formation	17
2.3. Image Restoration	23
2.4. Super-resolution	29
CHAPTER 3. MULTIREOLUTION METHODS IN IMAGE PROCESSING	33
3.1. Overview	33
3.2. Laplacian Pyramid	34
3.3. Wavelet transforms	38
3.3.1. Background	38
3.3.2. Approximation - Scaling coefficients	39
3.3.3. Detail - Wavelet Coefficients	43
3.3.4. The Discrete Wavelet Transform	47
3.4. Wavelet discussion - Frequency Domain	52
3.5. Variations on a theme	58
3.5.1. Biorthogonal Wavelet Transforms	59
3.5.2. Wavelet Packets	60
3.5.3. Undecimated Wavelet Transforms	61
CHAPTER 4. IMAGE PROCESSING WITH WAVELETS	64
4.1. Overview	64
4.2. Wavelet Denoising Algorithms	64
4.3. Image Restoration	72
4.4. Image Interpolation	75

TABLE OF CONTENTS—*Continued*

CHAPTER 5. RADIALLY SYMMETRIC TRANSFORMS	79
5.1. Overview	79
5.2. Motivation	79
5.3. Radially Symmetric Transforms	80
5.4. Modified mirror wavelet basis	89
5.5. Discussion of Image Restoration Algorithm	92
5.5.1. Algorithm overview	92
5.5.2. Noise variance estimation	93
5.5.3. Denoising approach	95
5.5.4. Wavelet basis	96
5.6. Results	98
5.7. Relationship to Curvelet Transforms	101
CHAPTER 6. SUPER-RESOLUTION METHODS	109
6.1. Overview of Wavelets and Super-resolution	109
6.2. Feasibility Experiments	111
6.2.1. Sparsity	112
6.2.2. Magnitude Extrapolation	116
6.2.3. Determination of Significance	117
6.2.4. Discussion	124
6.3. Vector Quantization	126
6.3.1. Overview	126
6.3.2. Experimental Results	128
CHAPTER 7. CONCLUDING REMARKS	141
REFERENCES	143

LIST OF FIGURES

FIGURE 2.1. Model of a general imaging system	18
FIGURE 2.2. 3-D plot of (a) $jinc^2$, the PSF for a circular aperture imaging system and (b) the related MTF.	21
FIGURE 2.3. 1-D plot of (a) $jinc$ function, (b) $jinc^2$ function, the PSF for a circular aperture imaging system, and (c) the related MTF. Note that the dotted line in (c) is the triangle function (straight line) for comparison.	22
FIGURE 2.4. 1-D plots of the resultant images from two point sources. Point sources are vertical lines, dotted lines are the PSF, and the solid line is the image. (a) is clearly resolved, (b) is the Rayleigh criteria, (c) is the Sparrow criteria, and (d) is non-resolved.	23
FIGURE 3.1. Examples of resolution in imagery: (a) is original Lenna image, (b) a high-resolution subband, and (c) a low-resolution subband.	33
FIGURE 3.2. Algorithm for Laplacian pyramid transform	35
FIGURE 3.3. Gaussian Pyramid of Lenna. (a) shows the original image and 2 levels of decomposition. (b) shows the same images as in (a), but upsampled to better visualize the effect of scale	36
FIGURE 3.4. Laplacian Pyramid of Lenna showing the first two Laplacian images (L_1 and L_2) and the resulting Gaussian image (G_2). (a) shows the downsampled images. (b) shows the same images as in (a), but upsampled to better visualize the effect of scale and displayed as absolute value to enhance large magnitude areas.	37
FIGURE 3.5. The Haar basis: (a) scaling function and (b) wavelet function	47
FIGURE 3.6. The Daubechies-4 and Daubechies-12 basis functions. (a) and (c) are for Daubechies-4, (b) and (d) are for Daubechies-12	48
FIGURE 3.7. The bi-orthogonal 9/7 basis functions. (a) and (c) form a pair; (b) and (d) are the other pair.	49
FIGURE 3.8. Algorithm for forward and inverse discrete wavelet transform	51
FIGURE 3.9. 2-D DWT display: (a) is location of subbands in display. (b) is DWT of Lenna image	52
FIGURE 3.10. Time-frequency distribution for a 1-D signal in the (a) time domain and (b) Fourier domain	53
FIGURE 3.11. Time-frequency distribution for 1-D signal using a STFT	54
FIGURE 3.12. Frequency response of Daubechies-12 filters. The scaling coefficient filter is solid, the wavelet coefficient filter is dashed.	55
FIGURE 3.13. Time-frequency distribution for a 1-D signal in the wavelet domain	57
FIGURE 3.14. Algorithm for forward and inverse discrete wavelet transform using bi-orthogonal wavelets	59

LIST OF FIGURES—Continued

FIGURE 3.15. Frequency response of bi-orthogonal 9/7 filters. The first pair of low-pass/high-pass filters is solid, the second pair is dashed.	60
FIGURE 3.16. (a) Frequency response of wavelet filters using the <i>à trous</i> algorithm. Solid lines are the first level, dashed is second level and dotted is third level in the decomposition. (b) is the resultant band-pass nature of the respective levels for an input with flat frequency content.	63
FIGURE 4.1. Algorithm for wavelet domain processing	64
FIGURE 4.2. Frequency bands for mirror wavelet basis (dashed lines) and inverse filter for diffraction limited imaging system (solid line)	73
FIGURE 4.3. Algorithm for mirror wavelet basis decomposition	74
FIGURE 5.1. (a) 2-D frequency domain tiling of separable mirror wavelet basis, positive frequencies quadrant only, (b) with diffraction cut-off of circular aperture imaging system	80
FIGURE 5.2. Frequency response of inverse filter for a circular aperture imaging system (solid line) along (a) either frequency axis and (b) the diagonal. Dotted lines denote the frequency bands of the mirror wavelet coefficients.	81
FIGURE 5.3. The masks used for the (left to right) LL, HL, LH, HH subbands	83
FIGURE 5.4. Top row is the first three wavelet masks and the resulting scaling function frequency mask for the <i>sinc</i> wavelet. Bottom row shows the corresponding wavelet/scaling images.	84
FIGURE 5.5. Top row is the first three wavelet masks and the resulting scaling function frequency mask. Bottom row shows the corresponding wavelet/scaling images.	86
FIGURE 5.6. Top row is the first three wavelet masks and the resulting scaling function frequency mask for the bi-orthogonal 9/7 wavelet pair. Middle row shows the same masks with applicable periodic lobes. Bottom row shows the corresponding wavelet/scaling images.	87
FIGURE 5.7. 1-D plot across 2-D kernel of HH band (top) and log plot to show decay (bottom). The x-axis is pixels, while the y-axis is relative amplitude.	90
FIGURE 5.8. Plots for measurement of sparsity of transform representations comparing RS-DWT and undecimated DWT. In top plot RS-DWT is solid and undecimated DWT is dotted.	90
FIGURE 5.9. Plots for measurement of sparsity of transform representations comparing RS-DWT and ideal band-pass image of same frequency band. In top plot RS-DWT is solid.	91
FIGURE 5.10. Correlation coefficient of wavelet coefficients for RS-DWT (solid line) and undecimated DWT (dotted line) plotted vs. pixel separation. For comparison, the correlation of a white Gaussian noise image is shown as a dashed line.	91

LIST OF FIGURES—*Continued*

FIGURE 5.11. Graphical algorithm for calculating the RS-DWT mirror basis transform.	92
FIGURE 5.12. Frequency plot of the subbands in a RS-DWT mirror basis. Note that only positive frequency quadrant is shown.	93
FIGURE 5.13. Algorithm overview for RS-DWT based deconvolution.	94
FIGURE 5.14. Wiener filter denoising algorithm	97
FIGURE 5.15. Images used in restoration. (a) is Lenna, (b) is Urban.	99
FIGURE 5.16. Restoration results. Enlarged section of image to show detail. (a) is original, (b) is degraded image (SNR 30.62), (c) results of fair WF (ISNR 5.37), (d) results using RS-DWT (ISNR 6.26), (e) results of the unfair WF (ISNR 6.45), (f) results using RS-DWT with wavelet domain WF (ISNR 6.78).	102
FIGURE 5.17. Restoration results. Enlarged section of image to show detail. (a) is original, (b) is degraded image (SNR 25.66), (c) results of fair WF (ISNR 2.85), (d) results using RS-DWT (ISNR 5.04), (e) results of the unfair WF (ISNR 5.82), (f) results using RS-DWT with wavelet domain WF (ISNR 5.04).	103
FIGURE 5.18. Restoration results. Enlarged section of image to show detail. (a) is original, (b) is degraded image (SNR 25.66), (c) results of fair WF (ISNR 2.85), (d) results using RS-DWT (ISNR 5.04), (e) results of the unfair WF (ISNR 5.82), (f) results using RS-DWT with wavelet domain WF (ISNR 5.04).	104
FIGURE 5.19. 1-D horizontal plots near the bottom (row 410 of 512) of the Urban image. The plots are offset from one another for visibility. From bottom to top, they are the original image, the degraded image, the fair WF estimate, the unfair WF estimate, and the RS-DWT estimate.	105
FIGURE 5.20. Restoration results. Enlarged section of image to show detail. (a) is original, (b) is degraded image (SNR 24.91), (c) results of fair WF (ISNR -0.36), (d) results using RS-DWT (ISNR 1.87), (e) results of the unfair WF (ISNR 2.63), (f) results using RS-DWT with wavelet domain WF (ISNR 2.06).	106
FIGURE 5.21. Restoration results. Enlarged section of image to show detail. (a) is original, (b) is degraded image (SNR 24.91), (c) results of fair WF (ISNR -0.36), (d) results using RS-DWT (ISNR 1.87), (e) results of the unfair WF (ISNR 2.63), (f) results using RS-DWT with wavelet domain WF (ISNR 2.06).	107
FIGURE 5.22. Frequency domain representations of (a) the directional filter bank, (b) a multiscale decomposition, (c) the PDFB with number of directions doubling at every other scale	108

LIST OF FIGURES—*Continued*

FIGURE 6.1. Overview of super-resolution process in the wavelet domain. . .	109
FIGURE 6.2. Three methods of wavelet based super-resolution: (a) uses the standard decimated DWT, (b) the undecimated transform, and (c) the RS-DWT.	110
FIGURE 6.3. Images used in these experiments	112
FIGURE 6.4. Interpolation results for 10% of coefficients and random perturbations on the wavelet coefficients.	114
FIGURE 6.5. Interpolation results for 1% of the coefficients and random perturbations on the wavelet coefficients.	115
FIGURE 6.6. Plot of average correlation coefficient vs. percentage of coefficients included for the use of actual magnitudes (solid), random perturbation (dashed), and mean value (dotted).	116
FIGURE 6.7. Correlation averages for Lenna (solid), Urban (dotted) and Mandrill (dashed) given random perturbation of the magnitudes.	117
FIGURE 6.8. Results of predicting new subband magnitude from exponential decay. Left to right are LH, HL, HH subbands. Top to bottom are Lena, Urban, and Mandrill images.	118
FIGURE 6.9. Results of wavelet super-resolution algorithm using known location, but predicted magnitude of 1% of the coefficients.	119
FIGURE 6.10. Parent-child relationship in a 2-D DWT	120
FIGURE 6.11. Parent-Child relationships for a 1-D signal for (a) Decimated and (b) Undecimated DWTs.	120
FIGURE 6.12. Parent-Child relationships for a 1-D signal for an Undecimated DWT.	121
FIGURE 6.13. Test image used in experiments	123
FIGURE 6.14. Plot of SSR and LR for 10% significance at scale j	124
FIGURE 6.15. Plot of SSR and LR for 5% significance at scale j	125
FIGURE 6.16. Results from VQ-based pass-band only restoration using separable DWT. (a) is original image, (b) is blurred image, SNR=21.50 dB, (c) is restored image, ISNR=3.28 dB, (d) is spectral correlation image. . . .	133
FIGURE 6.17. Graphical representation of 3 level transform for (a) separable DWT and (b) RS-DWT.	134
FIGURE 6.18. Results from VQ-based pass-band only restoration using RS-DWT. (a) is original image, (b) is blurred image, SNR=21.62 dB, (c) is restored image, ISNR=6.75 dB, (d) is spectral correlation image.	135
FIGURE 6.19. Results for super-resolution using separable DWT. (a) is original, (b) is blurred, SNR=17.11 dB, (c) is restored, ISNR=3.28 dB, (d) is spectral correlation plot.	136

LIST OF FIGURES—*Continued*

FIGURE 6.20. Results for super-resolution using RS-DWT uniform packet basis. (a) is original image, (b) is band-limited image SNR=29.67, (c) is restoration ISNR= -0.81 dB, and (d) is spectral correlation of restoration with original. The SR-ISNR is 0.15 dB.	137
FIGURE 6.21. Images of wavelet subbands for the NLIVQ estimation (left) and truth data (right) of the same subband which is beyond the band-limit (super-resolution).	138
FIGURE 6.22. 1-D plot of the wavelet coefficients. Top is the NLIVQ based estimation, bottom is the truth data.	138
FIGURE 6.23. Results for DCT based NLIVQ estimation. (a) is the original, (b) is band-pass image SNR=23.76, and (c) is the NLIVQ restoration, SNR=24.41. The SR-ISNR is 0.95 dB.	139
FIGURE 6.24. Log-compressed image of frequency content of true image (left) and NLIVQ estimate (right). Black circles show the band-pass cut-off. .	140
FIGURE 6.25. Complex correlation plot of NLIVQ estimation.	140

LIST OF TABLES

TABLE 3.1.	Filter coefficients for various wavelet bases	47
TABLE 5.1.	RS-DWT deconvolution results	100

ABSTRACT

Multi-resolution techniques, and especially the wavelet transform provide unique benefits in image representation and processing not otherwise possible. While wavelet applications in image compression and denoising have become extremely prevalent, their use in image restoration and super-resolution has not been exploited to the same degree. One issue is the extension 1-D wavelet transforms into 2-D via separable transforms versus the non-separability of typical circular aperture imaging systems. This mismatch leads to performance degradations.

Image restoration, the inverse problem to image formation is the first major focus of this research. A new multi-resolution transform is presented to improve performance. The transform is called a Radially Symmetric Discrete Wavelet-like Transform (RS-DWT) and is designed based on the non-separable blurring of the typical incoherent circular aperture imaging system. The results using this transform show marked improvement compared to other restoration algorithms both in Mean Square Error and visual appearance. Extensions to the general algorithm that further improve results are discussed.

The ability to super-resolve imagery using wavelet-domain techniques is the second major focus of this research. Super-resolution, the ability to reconstruct object information lost in the imaging process, has been an active research area for many years. Multiple experiments are presented which demonstrate the possibilities and problems associated with super-resolution in the wavelet-domain. Finally, super-resolution in the wavelet domain using Non-Linear Interpolative Vector Quantization is studied and the results of the algorithm are presented and discussed.

CHAPTER 1

INTRODUCTION

1.1 Problem Definition

The heart of this work is a simple goal that has been around for a long time: processing digital imagery to make it better. Obviously 'better' is a subjective term open to various interpretations. It also implies that the images we receive have been degraded in some manner. The assumption throughout this work is that the degradations stem from a circular aperture imaging system further corrupted by noise.

Even before computer-based techniques were feasible, people have sought methods to process images to mitigate the effects of the imaging system (e.g. blur, noise) or the environment (e.g. atmospheric distortion, dim objects). Over the years, this has grown into a large body of work with the majority of the current work focused on digital imagery. The approach I have taken is to look at image processing from a multi-resolution perspective. Wavelets, the mathematical basis of multi-resolution decomposition, have a long history in signal processing, and more recently have been extended to 2-D applications such as image processing. Using the unique attributes of multi-resolution decompositions, improved estimates of the imaged object are shown to be possible.

1.2 Research Goals

The primary goals of this research are

- Analyze image restoration from a multi-resolution perspective
- Develop new algorithms for image restoration to improve performance

- Investigate the feasibility of super-resolution in the wavelet domain.
- Analyze Vector Quantization as a means to achieve super-resolution.

1.3 Outline

Chapter 2 provides the necessary background on image formation. This starts with the most general case, and discusses the approximations that are necessary to simplify the problem. Next it discusses the inverse problem, recovering knowledge of an object given its degraded image. Discussion on the resolution of an imaging system and metrics used to quantify restoration performance are next. It concludes with a discussion of super-resolution and how the definition varies in slight but important ways throughout the literature.

Chapter 3 contains an introduction to the concept of multiresolution image decomposition. Starting with the Laplacian Pyramid, it discusses the motivations and techniques behind multiresolution image decompositions as well as their implementation. The wavelet transform is then derived from this basic concept of multiresolution decomposition. The properties of the transform are discussed in both the spatial and frequency domain. Last, several variations of the general transform that are required in later chapters are presented.

Chapter 4 briefly summarizes the large body of previous work in the field of image processing using wavelet transforms. It starts with a discussion on image denoising applications. Next, image restoration algorithms are presented and connected to image denoising. Last, a few image interpolation techniques are presented, which I'll later connect within the construct of super-resolution.

Chapter 5 introduces a new method of multiresolution decomposition that aligns more closely to the typical imaging systems in use. The transform will be described and applied to the image restoration problem. Results will be presented and compared with other techniques.

Chapter 6 will discuss super-resolution techniques from a multiresolution perspective. A set of experiments to demonstrate the feasibility are presented, along with a proposed algorithm to super-resolve based on Vector Quantization.

Finally, Chapter 7 offers a summary of results and a discussion of the areas that seem most promising for future study.

CHAPTER 2

FUNDAMENTALS OF IMAGE FORMATION AND RESTORATION

2.1 Overview

Any rationale attempt to restore imagery must be based on a reasonable model of how the image was formed from the object. Constructing such a model, along with the necessary simplifications to make it tractable, is the basis of the first section of this chapter. From this basis, we can begin to strategize on methodologies for the inverse problem – recovering knowledge of the object from the image. A discussion of some of the classical methods of image restoration is the focus of the second section. In addition, the concept of super-resolution, an extension of image restoration, is discussed in the last section.

The areas of image formation and restoration are a maturing field and have a considerable amount of literature dedicated to them. In this chapter, I seek to uncover parts of these topics which are critical to subsequent chapters rather than cover it in detail. For additional background and more extensive treatments in image formation, two classic texts are [1, 2] which develop the results from the basic theory. Image restoration is a broad discipline, but several excellent starting points for further study are [3, 4, 5, 6] as well as survey article [7] and references therein.

2.2 Image Formation

Image formation is the process of detecting radiant energy emanating from an object. This is graphically depicted in figure 2.1. Denoting the object by $f(x', y')$

and the system by an operator \mathcal{L} , we can write the detected image

$$g(x, y) = \mathcal{L} \{f(x', y')\} \quad (2.1)$$

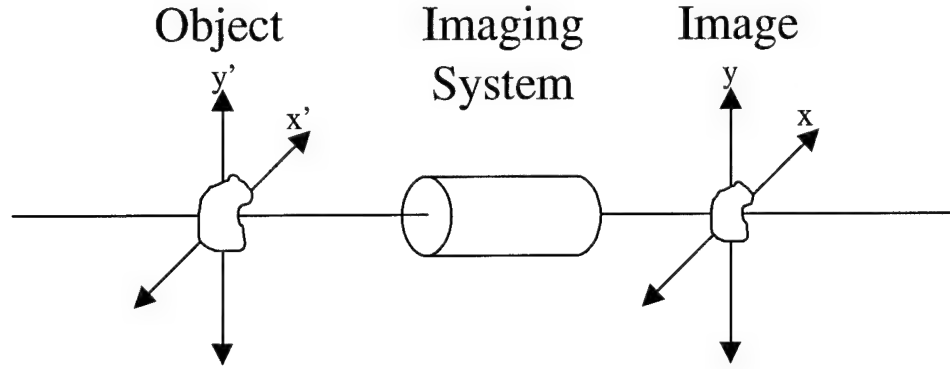


FIGURE 2.1. Model of a general imaging system

In general, the resultant image from the propagation of a monochromatic optical wave field from the object plane to the image plane can be written as

$$g(x, y) = \iint h(x, y, x', y', f(x', y')) dx' dy' \quad (2.2)$$

where h is the response at the point (x, y) in the image plane to an impulse of amplitude $f(x', y')$ at point (x', y') in the object plane, accounting for the optical wave propagation and system effects. A result of diffraction theory and the linear nature of diffraction shows that, in the Fraunhofer region, the image is given by

$$g(x, y) = \iint h(x, y, x', y') f(x', y') dx' dy' \quad (2.3)$$

where $h(x, y, x', y')$ is the amplitude response at the image point (x, y) to an impulse at (x', y') in the object plane. Here we make the assumption that wave propagation and the imaging system are linear which removes the functional dependence on the amplitude, $f(x', y')$, from h . Equation 2.3 defines a broad class of problems encountered in many physical situations, and is called a *Fredholm integral equation of the*

first kind. An attempt to estimate $f(x', y')$ from the data $g(x, y)$ is called an *inverse problem*.

At this point, we need to make an additional assumption about the system in order to simplify the problem. First, we assume the optical system has no aberrations, it is “in focus,” and that the object and image planes are stationary. Together, these imply that the system is *shift invariant* (SI) – that the amplitude in the image plane from a impulse in the object plane is dependent only on the relative separation, i.e. $h(x, y, x', y') = h(x - x', y - y')$. This leads to the model

$$g(x, y) = \int \int h(x - x', y - y') f(x', y') dx' dy' \quad (2.4a)$$

$$g(x, y) = (h \otimes f)(x, y) \quad (2.4b)$$

where $(h \otimes f)(x, y)$ denotes the convolution of h and f evaluated at (x, y) . Thus, for a *linear and shift-invariant* (LSI) system, the resultant image is a convolution of the object with h .

As this dissertation will be focused only on incoherent imaging, equation 2.4b requires further discussion. While g and f above are amplitudes, in the incoherent case we can only measure the time average intensity. Thus we need the relationship between the object and image intensities (g_i and f_i , respectively), which is given by

$$g_i(x, y) = (|h|^2 \otimes f_i)(x, y) \quad (2.5)$$

The image intensity is the convolution of the object intensity with the square of the amplitude impulse response. The $|h|^2$ term, the intensity impulse response, is called the *point-spread function* (PSF).

The PSF can be derived from the aperture function $p(x, y)$ which is the transmission of the system aperture. For an aberration free system, $p(x, y)$ is one inside the aperture and zero outside. From diffraction theory, the PSF is given by the magnitude squared of the Fourier transform of the aperture function (within a constant

multiplying scale factor) evaluated at frequencies $(\xi, \eta) = (\frac{x}{\lambda d}, \frac{y}{\lambda d})$

$$|h(x, y)|^2 \propto |\mathcal{F}\{p(x, y)\}|^2 \quad (2.6a)$$

$$\propto \left| P\left(\frac{x}{\lambda d}, \frac{y}{\lambda d}\right) \right|^2 \quad (2.6b)$$

As for notation, uppercase variables represent the Fourier domain representation of the function, $\mathcal{F}\{p(x, y)\} = P(\xi, \eta)$. For simplicity of notation, in subsequent sections I will drop the squared term and refer to the PSF as simply h . Additionally, I will drop the primed notation on the object plane coordinates unless needed for clarity.

If we look in the frequency domain, the Fourier transform of the PSF is called the *optical transfer function* (OTF), $H(\xi, \eta)$, and its magnitude is the *modulation transfer function* (MTF). The OTF is the frequency response of the system, and is by definition normalized to unity at the DC term.

$$H(\xi, \eta) = \mathcal{F}\{|h(x, y)|^2\} \quad (2.7a)$$

$$\propto p(x, y) \star p^*(x, y) \quad (2.7b)$$

where \star denotes autocorrelation. From this we see that the OTF is directly proportional to the autocorrelation of the aperture function.

Since the OTF of any real system is the autocorrelation of a (necessarily) finite aperture function, the OTF must have finite extent. This is critical result: any LSI system will eliminate all frequencies above a threshold, f_c , which is determined uniquely by the aperture function. This frequency above which no information can be passed by the system is called the *diffraction cut-off*. For a circular aperture of diameter d and focal length f , the MTF (in polar form) is given by [1, eqn 6-32]

$$H(\rho) = \begin{cases} \frac{2}{\pi} \left[\cos^{-1}\left(\frac{\rho}{\rho_0}\right) - \frac{\rho}{2\rho_0} \sqrt{1 - \left(\frac{\rho}{2\rho_0}\right)^2} \right] & \rho \leq 2\rho_0 \\ 0 & \text{otherwise} \end{cases} \quad (2.8)$$

where $\rho = \sqrt{\xi^2 + \eta^2}$ and $\rho_c = 2\rho_0 = \frac{2d}{\lambda f}$ is the cut-off frequency,. The corresponding

PSF (within a constant multiplying scale factor) is given by

$$h(r) \propto \text{jinc}^2(r) \propto \left(\frac{J_1(\pi r)}{\pi r} \right)^2 \quad (2.9)$$

where J_1 is the first order Bessel function of the first kind and $r = \frac{d}{\lambda f} \sqrt{x^2 + y^2}$. For the circular aperture case, 3-D plots of the MTF and PSF are shown in figure 2.2, while a 2-D slice of the MTF and PSF through the maximum is shown in figure 2.3. I will refer to this PSF and OTF as the incoherent circular aperture case.

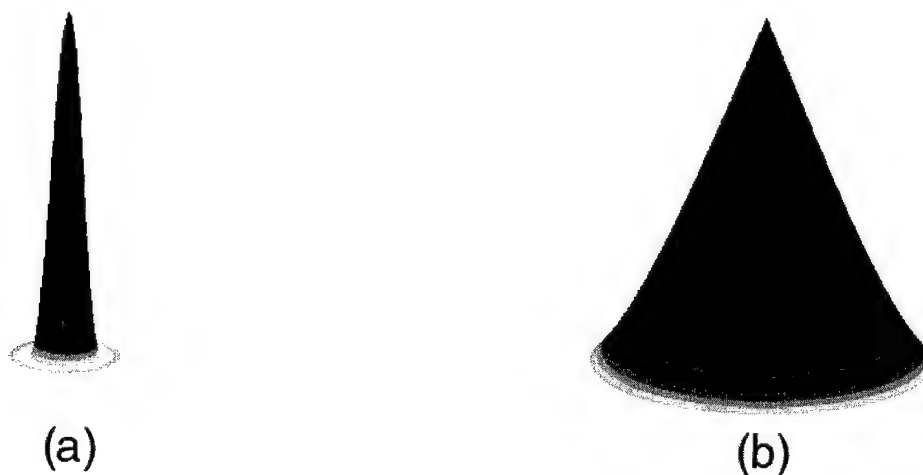


FIGURE 2.2. 3-D plot of (a) jinc^2 , the PSF for a circular aperture imaging system and (b) the related MTF.

One item of primary interest is the resolution of a given image. This term is subjective in its nature¹, but criteria can be established to permit specification and comparison of system resolutions. The results are derived for a circular aperture imaging system sampled at or above the Nyquist rate. Four separate cases for images of two point sources are shown in figure 2.4. Due to the circular symmetry of the PSF, an arbitrary 1-D plot through the center is sufficient. The first plot (figure

¹Lord Rayleigh, who developed the first criteria discussed below, stated "This rule is convenient on account of its simplicity and it is sufficiently accurate in view of the necessary uncertainty as to what exactly is meant by resolution" [8]

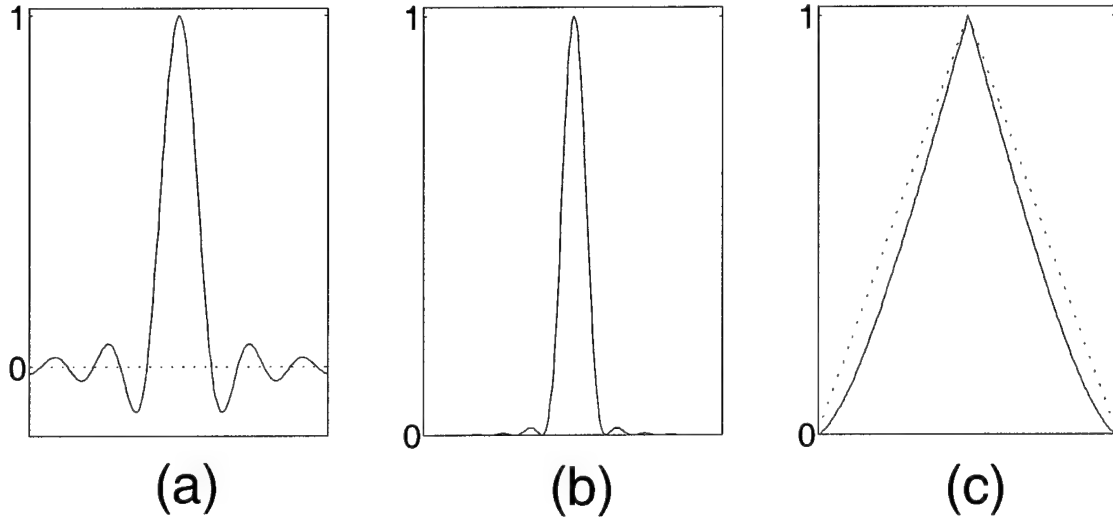


FIGURE 2.3. 1-D plot of (a) *jinc* function, (b) *jinc*² function, the PSF for a circular aperture imaging system, and (c) the related MTF. Note that the dotted line in (c) is the triangle function (straight line) for comparison.

2.4a) shows when the two point sources are clearly resolved. Next is what is termed the *Rayleigh criteria* - when the first zero in the PSF corresponds to the maximum in the other. The next is *Sparrow's criteria* when the resultant image has flat midpoint between the objects, or more precisely when the second derivative is zero. The last is when the two points are clearly not resolved. For a non-circularly symmetric system, similar resolution criteria will depend on orientation.

The Rayleigh criteria is most broadly known as the resolution limit of the system, and is when the point sources are separated by a distance

$$x_{sep} = 1.22 \frac{\lambda r}{d} \quad (2.10)$$

where r is the range to the object. In actuality, this criteria is pessimistic as to what can actually be resolved[8], especially including restoration techniques discussed in the next section. The Sparrow's criteria may be closer to reality, but leads to a more complicated derivation of the resolution distance.

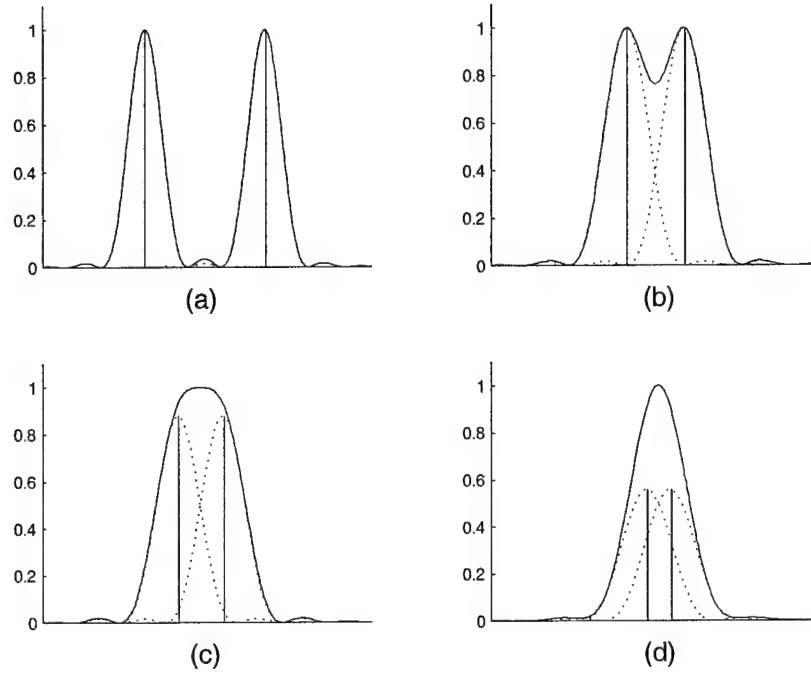


FIGURE 2.4. 1-D plots of the resultant images from two point sources. Point sources are vertical lines, dotted lines are the PSF, and the solid line is the image. (a) is clearly resolved, (b) is the Rayleigh criteria, (c) is the Sparrow criteria, and (d) is non-resolved.

2.3 Image Restoration

Image restoration is the attempt to reconstruct the original object, $f(x, y)$ from the data $g(x, y)$. The estimate of the object will be denoted with an overhat, \hat{f} . The above discussion on image formation is based on mapping a continuous function (object) to another continuous function (image), $\mathcal{L}_{cc} : f(x, y) \rightarrow g(x, y)$. While reasonable in theory, actual implementations use digital data which is by its essence discrete. A more practical point of view is the *continuous-to-discrete* mapping, $\mathcal{L}_{cd} : f(x, y) \rightarrow g[m, n]$, where the object plane is allowed to remain continuous, while the image is discrete - composed of individual pixel values indexed by integers m and n :

$$g[m, n] = \int \int h_{mn}(x, y) f(x, y) dx dy \quad (2.11)$$

where $h_{mn}(x, y)$ is the analog to the continuous PSF - i.e. the sensitivity of the $[m, n]^{th}$ pixel to light from (x, y) in the object plane. This mapping represents the reality of the typical imaging system, where the object is continuous, but the data is discrete. However, there will be infinitely many continuous objects that produce a single discrete image, due to the change in dimensionality of the problem.

This leads to the concept of a null space. The null space is a function of the PSF and is defined by

$$S_f = \left\{ \hat{f}(x) : \left(\hat{f} \otimes h \right) (x, y) - g(x, y) = 0 \right\} \quad (2.12)$$

which essentially states that any estimate that matches the data is a member of the null space. A null object is defined as the difference between a member of the null space and the true object. Obviously, any image with content only above the diffraction cut-off will be a null object. Other null objects are possible due to the continuous nature of the object and discrete nature of the image. The best we can accomplish is to obtain an estimate within the null space meeting some regularization constraint.

In digital implementation of image restoration, the object estimate is also limited to discrete approximations, although this may be displayed in a continuous form with the use of suitable basis function expansion. Thus, in order to readily compute performance metrics between the original and restored images, a *discrete-to-discrete* mapping is useful, $\mathcal{L}_{dd} : f[m, n] \rightarrow g[m, n]$

$$g[m, n] = \sum_{m'} \sum_{n'} h[m, n, m', n'] f[m', n'] \quad (2.13)$$

where h again is the analog to a PSF, the sensitivity of the $[m, n]^{th}$ pixel to light from point $[m', n']$ in the object plane. This perspective permits use of digital image data and allows comparison of the estimation of the object, $\hat{f}[m, n]$, to the original discrete representation of the object. However, this is only an approximation of what is truly happening: a continuous object being mapped into a discrete image.

It will be assumed throughout the dissertation that the discrete image is sampled above the Nyquist rate to prevent any aliasing, i.e. samples are closer than $\frac{1}{2f_{max}}$ for a bandlimited signal with maximum frequency f_{max} .

As discussed in the above section, the frequency response of any system will be non-zero only out to a cut-off frequency f_c . Image restoration is the attempt to restore the frequency content of the original image. For the purposes of this discussion, I will make the distinction of frequency content below the diffraction cut-off, a concern of classical image restoration (discussed below in this section), and that above the diffraction cut-off, which is restored through super-resolution techniques to be discussed in the next section. In practice, however, they are often combined into a single algorithm.

Let us take as a model the typical linear shift-invariant system:

$$g(x, y) = (h \otimes f)(x, y) + n(x, y) \quad (2.14)$$

where n is the noise. In the Fourier domain we have

$$G(\xi, \eta) = H(\xi, \eta) F(\xi, \eta) + N(\xi, \eta) \quad (2.15)$$

The straightforward approach of inverting H , (dropping the indices for clarity)

$$\hat{F} = H^{-1}G = F + H^{-1}N \quad (2.16)$$

is often unstable due to any zeros (or low values) in the system frequency response H which will amplify the noise. This is usually a useless result. However, we can see from this representation that the inverse filtered estimate is simply the original object with noise $H^{-1}N$. Thus, the deconvolution is now a denoising problem. However, even for additive white Gaussian noise (AWGN), $H^{-1}N$ will not be white. The inverse filter will color the noise, often significantly.

An initial approach to deal with this is the use of a object consistency metric where minimization of a distance metric between the estimate and true object is desired.

$$\hat{f} = \underset{\hat{f}}{argmin} \left\{ d(\hat{f}, f) \right\} \quad (2.17)$$

Of course, the true object is rarely known which requires the use of a data consistency metric

$$\hat{f} = \underset{\hat{f}}{\operatorname{argmin}} \left\{ d \left(h \otimes \hat{f}, g \right) \right\} = \underset{\hat{f}}{\operatorname{argmin}} \left\{ Q_{data} \left(\hat{f}, g \right) \right\} \quad (2.18)$$

One of the most well-known restoration methods, the *Wiener filter*, is an example of this, where the desire is to minimize the mean-square error,

$$\hat{f} = \underset{\hat{f}}{\operatorname{argmin}} \left\{ \sum_{x,y} \left| f(x,y) - \hat{f}(x,y) \right|^2 \right\} \quad (2.19)$$

consistent with AWGN n , and the fact that the estimate will be a linear filtered version of the data, $\hat{f} = w \otimes g$. The resultant filter in the frequency domain is

$$W(\xi, \eta) = \frac{H^*(\xi, \eta)}{|H(\xi, \eta)|^2 + \frac{S_n(\xi, \eta)}{S_f(\xi, \eta)}} \quad (2.20)$$

$$\text{where } S_n(\xi, \eta) = \langle |N(\xi, \eta)|^2 \rangle \quad \text{and} \quad S_f(\xi, \eta) = \langle |F(\xi, \eta)|^2 \rangle$$

Thus the filter can be looked at as an inverse filter followed by reduction of the frequency magnitude based on the SNR $\left(\frac{S_n(\xi, \eta)}{S_f(\xi, \eta)} \right)$:

$$\hat{F} = WG \quad (2.21a)$$

$$= \left(\frac{1}{1 + \alpha} \right) H^{-1} G \quad (2.21b)$$

$$\text{where } \alpha = \left(\frac{S_n(\xi, \eta)}{S_f(\xi, \eta)} \cdot \frac{1}{|H(\xi, \eta)|^2} \right)$$

From Equation 2.21b, any frequency component will be multiplied by the inverse filter and then ‘shrunk’ in relation to a weighted SNR. For large SNR, $\left(\frac{1}{1+\alpha} \right)$ approaches unity maintaining the frequency content, while for small SNR, $\left(\frac{1}{1+\alpha} \right)$ approaches zero, shrinking its contribution to the restored image.

One issue in using the Wiener filter is the calculation of the *power spectral densities* (PSDs) $S_f(\xi, \eta)$ and $S_n(\xi, \eta)$. Techniques exist for approximating $S_n(\xi, \eta)$ from the data, especially in the white noise case. However, $S_f(\xi, \eta)$ is rarely well-known for any problem of interest. One solution is to estimate $S_f(\xi, \eta)$ iteratively from the data, see [9] for an example.

Any LSI restoration filter can be written as a multiplication in the Fourier domain. If W is an arbitrary LSI filter,

$$\hat{F} = WG = WHF + WN \quad (2.22a)$$

$$= F + (WH - I)F + WN \quad (2.22b)$$

$$= F + E_f + E_n \quad (2.22c)$$

Thus, there are two error terms in the restoration, E_f which is proportional to the object, and E_n which is proportional to the noise. Any regularization of an LSI filter is an attempt to balance the two terms. To restore the frequency content at a point where $H(\xi, \eta)$ is close to zero, we must increase $W(\xi, \eta)$ to compensate. This will also amplify the noise (E_n), which can cause more harm than good. If we reduce W to reduce noise amplification, the product WH will decrease and $(WH - 1)$ will tend toward -1 at those frequencies, yielding $\hat{F} = F - F + E_n = E_n$. This will blur the image and cause ringing due to removal of frequency content. This is one motivation for non-LSI (either spatially variant or non-linear) methods – to provide a more robust trade-off between noise suppression, blurring and ringing.

The choice of the Euclidean metric (equation 2.19) is arbitrary and others have been proposed. As an example, the Kullback-Leibler distance (actually a generalized distance metric) may be more applicable in Poisson noise cases.

A second broad class of algorithms use a functional minimization of the form

$$\hat{f} = \underset{\hat{f}}{\operatorname{argmin}} \left\{ Q_{data}(\hat{f}, g) + \mu Q_{reg}(\hat{f}) \right\} \quad (2.23)$$

where in addition to the data consistency term, Q_{data} , a regularization term, Q_{reg} is added to impose additional constraints. Examples of Q_{reg} are minimum norm estimates, maximum entropy, and maximum likelihood. The μ hyper-parameter, the weighting of the terms in importance, must also be defined. The solution to equation 2.23 is rarely simple to determine and often requires iterative techniques.

One other class of image restoration techniques is projection onto convex sets (POCS). Often, known constraints on the image form a convex set in solution space. If there are two or more such constraints, an object estimate can be determined by repeated projection onto the known sets,

$$\hat{f}_{n+1} = \mathcal{P}_{s_1} \mathcal{P}_{s_2} \cdots \mathcal{P}_{s_k} \hat{f}_n \quad (2.24)$$

where $\mathcal{P}_{s_k} \hat{f}_n$ is the projection of \hat{f}_n onto the set s_k . For convex sets, repeated projections will converge to an estimate that will depend on the intersection of the k sets. This will be either a unique solution (single point of intersection), the least-squares solution (no intersection), or one of many solutions (intersection is a set itself). This is a powerful and often straightforward technique. However, it requires *a priori* knowledge of what sets are applicable and can be extremely slow in convergence.

After an estimate is calculated, there is a need to quantify the performance of the image restoration process. It is important to remember that whenever ‘performance’ is discussed, it is not a simple matter due to the variety of applications for the processed imagery. Thus, in any comparison one must account for the end-use of the imagery, as further discussed in [6]. For the purposes of this work, a subjective improvement in visual appearance is desired. While valid as a goal, it lacks an objective metric. In order to quantify ‘goodness’ of a restoration, I will use some metrics that have been previously developed and used. The first is mean square error (MSE) of an image N pixels square

$$MSE = \frac{1}{N^2} \sum_{(x,y)} |f(x,y) - \hat{f}(x,y)|^2 \quad (2.25a)$$

$$= \left\langle |f - \hat{f}|^2 \right\rangle \quad (2.25b)$$

The MSE can be compared to the mean value of the image to account for differences in magnitude of images. When calculated in *dB*, this is the *image SNR*. This can be used to compare any two images, whether the original and restored images, or the

original and degraded images:

$$SNR_{dB} = 10 \log_{10} \left(\frac{\langle |f|^2 \rangle}{\langle |f - \hat{f}|^2 \rangle} \right) \quad (2.26)$$

This can also be normalized to the peak value of f rather than the expectation, which is called the *peak SNR* (PSNR)

$$PSNR_{dB} = 10 \log_{10} \left(\frac{\max_{(x,y)} |f(x,y)|^2}{\langle |f - \hat{f}|^2 \rangle} \right) \quad (2.27)$$

To measure the image degradation from the combined effects of blur and noise, the *blurred SNR* (BSNR) can be used

$$BSNR_{dB} = 10 \log_{10} \left(\frac{\langle |(h \otimes f) - \langle (h \otimes f) \rangle|^2 \rangle}{\sigma_n^2} \right) \quad (2.28)$$

The *improvement SNR* (ISNR) is a metric that measures the improvement of the restored image, \hat{f} , compared to both the original, f , and degraded image, g .

$$ISNR_{dB} = 10 \log_{10} \left(\frac{\langle |f - g|^2 \rangle}{\langle |f - \hat{f}|^2 \rangle} \right) \quad (2.29)$$

Each of the above metrics is of some use in comparing images, but all suffer the same problem: there is no agreed upon global metric that can capture the performance of image restoration. I would add that there probably never will be one due to the variety of applications and desires. In the results presented later, I will include the above metrics, but also the images themselves for a subjective comparison.

2.4 Super-resolution

As discussed above, diffraction imposes a cut-off beyond which the frequency content of the object is lost. Super-resolution is the attempt to recreate some of this ‘lost’ frequency content. A working definition of super-resolution for my purposes is:

"The meaningful restoration of the object frequency spectrum for frequencies beyond the optical system cut-off."

Thus there are two important criteria for declaring super-resolved imagery: 1) there is restoration above the diffraction cut-off and 2) this restoration is meaningful.

Lacking an approved standard definition, super-resolution has been defined in similar, but different terms. Thus it is important to gain an understanding of the differences so we compare only apples to apples. For example, in [10], super-resolution is achieved using a wavelet-based interpolation algorithm. Using multiple sub-pixel shifted images of the same scene a zoom factor of 4 is achieved. This is far above the improvement reported elsewhere[11, 12]. However, there is a critical difference of definition. While the definition I use above is based on recovering information beyond the optical system cut-off, the implied definition for super-resolution in [10] is restoration of frequency spectrum above the detector array cut-off. In this case, the optics is not the limiting factor, but the detector is undersampled by a large amount. Thus, the image information contained in the higher frequencies (up to the optical system cut-off) is present in the imagery, albeit in aliased form. Thus the problem is to un-alias the frequency content using multiple realizations of the image with sub-pixel shifts – trading off temporal bandwidth for spatial bandwidth. This is a vastly different problem than that of actually restoring frequencies beyond the system cut-off, about which all *direct* information is eliminated. For the purposes of this research and the remainder of this dissertation, super-resolution will refer to the definition above, meaning the ability to restore frequency content lost due to the optical system cut-off and not due to undersampling.

The fundamental reason super-resolution is possible is based on the analytic continuation theorem. The mathematical theorems state that 1) any spatially bounded function will have an analytic Fourier transform and 2) for any analytic function known exactly in a finite region, the entire function can be uniquely determined. Any

spatially bounded object will have an analytic Fourier transform. Perfect knowledge of this in any finite region, will allow us to determine it everywhere - thus know its spectrum beyond the diffraction cut-off. While this lays out a mathematical explanation for why super-resolution may be possible (at least for spatially compact objects), it lays out untenable requirements. To know the Fourier transform exactly within a finite region requires no noise in the measurements. However, noise is always present, even if simply the quantization noise of digital systems. The super-resolution problem is very ill-conditioned, and becomes impossible as noise increases. However, there has been success in super-resolving imagery.

As discussed above, LSI systems can be completely described by a convolution in the spatial domain, or equivalently a multiplication in the frequency domain. Thus, an LSI restoration algorithm cannot super-resolve since a frequency content of zero can never be made non-zero through multiplication with the (finite) filter frequency response. Two aspects that are involved in successful super-resolution algorithms are:

1. The use of *non-linear* and/or *spatially varying* techniques.
2. The use of *a priori knowledge* such as positivity, limited spatial extent, distribution model etc. to regularize the problem

Multiple techniques to achieve super-resolution have been developed. The first discussion of super-resolution was in 1955[13]. A few techniques are based on a spatially varying algorithm, such as the Gerchberg-Papoulis algorithm[14, 15]. Most, however, are nonlinear and have been derived from Bayesian or maximum entropy estimates[16, 17, 18, 19, 20, 21]. A thorough discussion with historical background is available in either [22, 23]. One topic of interest here is how to quantify or measure super-resolution. One obvious method is to visually compare the 'super-resolved' image to the original. This is simple, but again lacks a quantitative metric. In

order to quantify the effect, I will use a *complex correlation coefficient image* (CCCI) as developed by Miller[24]. The idea is to measure the correlation of the (complex) Fourier domain coefficients of the original and super-resolved images by

$$\gamma(\xi, \eta) = \frac{\sum_{(u,v) \in N_k} F(u, v) \hat{F}^*(u, v)}{\left(\sum_{(u,v) \in N_k} |F(u, v)|^2 \right) \left(\sum_{(u,v) \in N_k} |\hat{F}(u, v)|^2 \right)} \quad (2.30)$$

where $N_k = \{(u, v) : \xi - k \leq u \leq \xi + k, \eta - k \leq v \leq \eta + k\}$

This is based on the correlation of random processes[25, eqn 10-26]. Equation 2.30 calculates the correlation of the frequency content of the images within a $2k + 1$ size square box around the pixel of interest. In this work, $k = 3$ was used. The correlation will be in the range $[0, 1]$. A zero implying no similarity, one implying identical images. When viewed as an image, these values will show where there is strong correlation to the original data. For a simulated image, where the diffraction cut-off was below the highest image frequency, any strong correlation above the diffraction cut-off would imply some measure of super-resolution. This is then a measure not only if high-frequency content has been added to the degraded image, but whether the frequency content matches the original and is therefore meaningful. If a single data point is desired, this correlation can be averaged for all frequencies above the diffraction cut-off.

CHAPTER 3

MULTIRESOLUTION METHODS IN IMAGE PROCESSING

3.1 Overview

Multiresolution in its simplest form is the decomposition of an image into multiple images of different 'resolutions.' The concept of resolution is a measure of the relative size of image content that is contained in the image. Thus a low resolution can be viewed as a blurred version of the original, where only coarse details are present. High resolution contains the fine details of the image. Figure 3.1 shows the Lenna image, along with a high-resolution subband and a low-resolution subband. This concept of resolution and fine or coarse detail can be related to the Fourier domain, where high/low resolution corresponds to high/low frequency content – the exact relationship will be explained in more detail below. Through the initial work of Hubel and Wiesel[26] and later Daugmann[27], the human visual system appears to use a multiresolution, wavelet-type method for image processing[28].

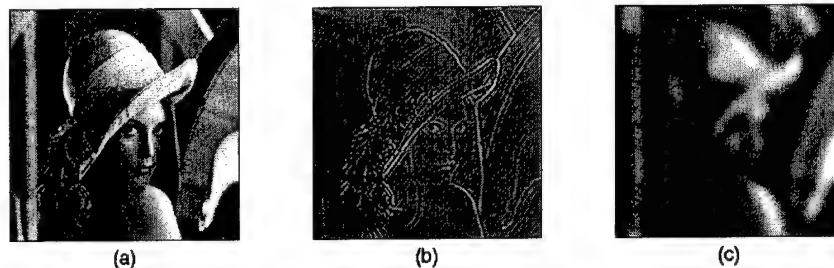


FIGURE 3.1. Examples of resolution in imagery: (a) is original Lenna image, (b) a high-resolution subband, and (c) a low-resolution subband.

This chapter will lay the theoretical and intuitive groundwork for the work presented in later chapters. Multiresolution techniques in image processing began in

earnest in 1983 with the Laplacian Pyramid of Burt and Adelson[29]. The first section is an overview of this work, demonstrating the overarching concept of multiresolution decomposition. In the late 1980's, this concept was combined with the wavelet transform[30] as it developed and matured. The wavelet transform has a long history before its current incarnation, with roots in several independent areas from geology to signal processing. The remainder of the chapter will be devoted to an explanation of the wavelet decomposition from both the spatial and frequency domain perspective. Compared to Laplacian pyramid based approaches, wavelets provide more flexibility in their construction, leading to many performance improvements.

To help clarify terminology, the use of 'signal' and 'image' will be used interchangeably, with an image simply being a 2-D signal. Likewise, in developing some of the concepts initially in 1-D, I will use the time domain interchangeably with the spatial domain. Last, the terms high-resolution, high-scale and detail representation are synonymous; similarly, low-resolution, small-scale and approximate representation are synonymous.

3.2 Laplacian Pyramid

The first significant application of multiresolution techniques in image processing was the Laplacian pyramid[29]. The algorithm used in this image transform is outlined in figure 3.2. A Gaussian pyramid is formed by low-pass filtering the original image and downsampling by a factor of 2 – they term this a REDUCE operation which is 'R' in the figure. This process can be repeated as desired to form successive levels, G_n , each a factor of 2 smaller both dimensions, as shown in figure 3.3. The term pyramid comes from the shape of the images when placed next to each other, and the Gaussian term comes from the limiting shape of the low-pass filter used. For 2-D images, it is assumed that the filters are applied in a separable manner, i.e. the same 1-D filter is applied to the rows, and then the columns to achieve the REDUCE

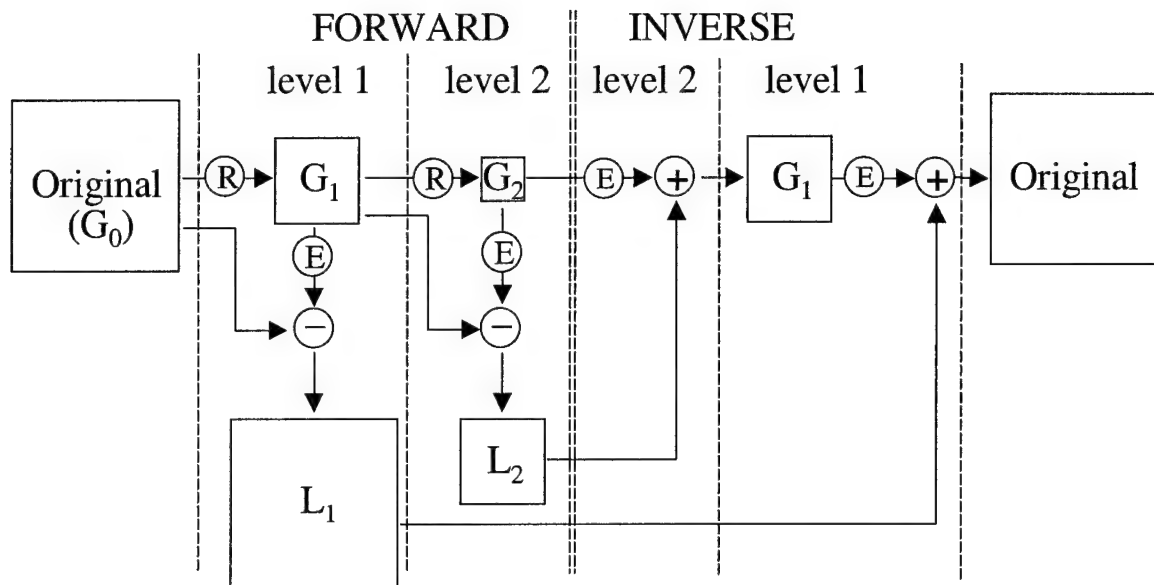


FIGURE 3.2. Algorithm for Laplacian pyramid transform

operation on a 2-D image.

To return the image to the original size, we can interpolate (termed EXPAND or 'E' in figure 3.2) the low-pass image. Due to the low-pass operator and downsampling, the expanded image will not be the same as the original. However, we can form an 'error' image by subtracting this from the original. This image, L_n , contains information as to how the low-pass version differs from the original. The error images are called a Laplacian Pyramid as the output is visually similar to that of a Laplacian operator. The complete transform consists of as many Laplacian images as desired (limited by image size) and a single resulting Gaussian image. For simplicity, I'll assume $n \times n$ images, where n is a power of 2. Thus, the maximum number of levels of decomposition is $\log_2 n$, where the resulting Gaussian image is a single pixel. The number of levels of decomposition actually computed is dependent on the application, but typically the decomposition is stopped when the Gaussian image is around 16 or 32 pixels square.

Two levels of the Laplacian pyramid (L_1 and L_2) and the remaining Gaussian im-

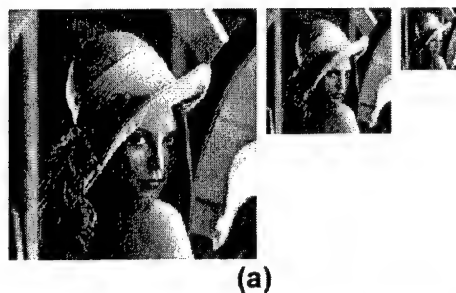


FIGURE 3.3. Gaussian Pyramid of Lenna. (a) shows the original image and 2 levels of decomposition. (b) shows the same images as in (a), but upscaled to better visualize the effect of scale

age (G_2) are shown in figure 3.4. The inverse transform is straightforward: given the Laplacian pyramid images and the base Gaussian image, we interpolate the Gaussian image, add it to the Laplacian image of the same level, and use the result as the Gaussian image at the next level. See figure 3.2 for an algorithmic outline. This method will provide a perfect reconstruction of the original image from the transform data. As is obvious from figure 3.3, successive Gaussian pyramid images are more blurred versions of each other - this becomes more obvious when the Gaussian images are interpolated to the original size. The first Laplacian pyramid image is a high-pass version of the original. Subsequent Laplacian images are similar to band-pass filtered images, while the last Gaussian image is a low-pass image. The smallest image details are visible in the L_1 image, notably the fine texture in the hair and the hat (see figure 3.4(b)), whereas L_2 shows prominence along the coarser texture of the hair and face. Note that a step-edge such as the hat outline will appear at multiple

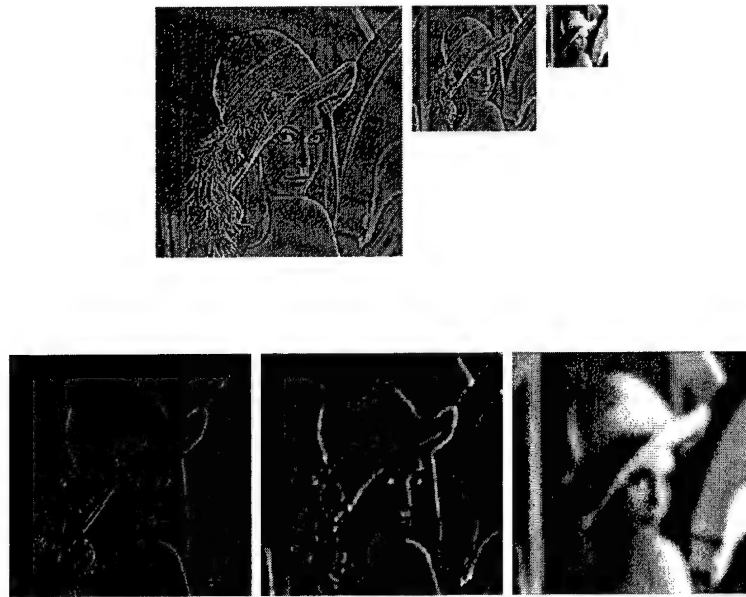


FIGURE 3.4. Laplacian Pyramid of Lenna showing the first two Laplacian images (L_1 and L_2) and the resulting Gaussian image (G_2). (a) shows the downsampled images. (b) shows the same images as in (a), but upsampled to better visualize the effect of scale and displayed as absolute value to enhance large magnitude areas.

levels.

One result of the algorithm is that there is a redundancy in the transform. The output will be a series of images, the first with dimensions of the original, the second half the size in each dimension, etc. Thus the number of output pixels will be greater than the number of input pixels, which means we have a redundancy in the data. The redundancy factor will depend on the number of levels of transform, but is bounded above by $\frac{4}{3}$ ($= 1 + \frac{1}{4} + \frac{1}{16} + \dots$). While this redundancy can be beneficial in some applications, in others such as compression it is a drawback. Another aspect of the transform to note is aliasing occurs in the downsampling operation. As with any low-pass FIR filter, when the result is downsampled aliasing will occur. While this is not an issue for the Laplacian pyramid in terms of perfect reconstruction due to the construction of the algorithm, it can cause trouble in other applications where

the transformed images are modified.

As demonstrated in the original paper, image compression performance was significantly improved, even though they start with a redundant representation of the data. The pyramidal form also allows for progressive transmission of the image - the Gaussian image is used as an initial estimate, followed by improvements as the Laplacian images are received. It was also proposed for computer vision applications since image features of various sizes (i.e. resolutions) are isolated in each of the Laplacian pyramid images.

3.3 Wavelet transforms

3.3.1 Background

This section is meant to introduce some mathematical concepts that will be used in subsequent sections. For a more thorough discussion see [31] and the extensive bibliography therein. The first definition is a *function space* which is a linear (infinite dimensional) vector space where the vectors are allowed to be continuous functions. An example is $L^2(\mathbb{R})$ which is the space of all square-integrable functions, i.e. $L^2(\mathbb{R}) = \left\{ f(x) : \int_{-\infty}^{\infty} |f(x)|^2 dx < \infty \right\}$. A set of vectors $\{\phi_k\}$ span a space \mathcal{S} if any element of \mathcal{S} can be written as a linear combination of the ϕ_k , i.e. for any $f(x) \in \mathcal{S}$,

$$f(x) = \sum_k a[k] \phi_k(x) \quad (3.1)$$

The set $\{\phi_k\}$ is then called an expansion set for \mathcal{S} . Note that the limits on the K summation may be infinite. If the expansion coefficients, $a[k]$, are unique, it is a basis set. If $\langle \phi_k, \phi_l \rangle = 0$ for $k \neq l$ then it is an orthogonal basis. If $\langle \phi_k, \phi_l \rangle = \delta(k - l)$, it is an orthonormal basis set. For an orthogonal basis, the expansion coefficients can be found by an inner product $a[k] = \langle f(x), \phi_k(x) \rangle$.

If, however, the ϕ_k 's are not independent, they can still form a frame. A frame is an overcomplete set of vectors (functions) that span the given set. To be a frame,

ϕ_k must satisfy

$$A|f|^2 \leq \sum_k |\langle f, \phi_k \rangle|^2 \leq B|f|^2 \quad (3.2)$$

for some $0 < A \leq B < \infty$ and all $f \in \mathcal{S}$. If $A = B$, then it is called a tight frame. This is a generalized Parseval's theorem, when $A = B = 1$ it reduces to Parseval's equality. The overcompleteness of a frame provides a redundancy in the expansion. In certain applications this can provide a more robust representation, while in others such as compression, it is not desired.

While orthogonal bases are simpler, at times a biorthogonal system will be desirable. In a biorthogonal basis, there are two sets, $\{\phi_k\}$ and $\{\check{\phi}_k\}$. The elements of each set are not orthogonal to each other, but to all members of the other set,

$$\langle \phi_k, \check{\phi}_l \rangle = \delta(k - l) \quad (3.3)$$

In a biorthogonal set, the vectors used to calculate the coefficients are different from those used in the expansion, leading to two possible decompositions:

$$f(x) = \sum_k a[k] \phi_k(x) = \sum_k \langle f(x), \phi_k(x) \rangle \check{\phi}_k(x) \quad (3.4a)$$

$$= \sum_k \check{a}[k] \check{\phi}_k(x) = \sum_k \langle f(x), \check{\phi}_k(x) \rangle \phi_k(x) \quad (3.4b)$$

The use of different sets for signal analysis and synthesis provides more freedom in their selection, a fact that will be important later.

3.3.2 Approximation - Scaling coefficients

Following a similar idea as the Laplacian pyramid transform, wavelets seek an alternative representation of the image in a multiresolution form. First, let us examine the approximation of a signal at a given resolution level. For simplicity, the initial focus will be real, continuous functions, with the discrete case discussed later. Any function, $f(x) \in L^2(\mathbb{R})$ can be expanded in terms of a (possibly countably infinite) set of basis functions ϕ_k , within certain restrictions not important to this discussion.

Let us determine the function space spanned by translates of a single basis function, ϕ_0 , assuming a limited spatial extent. As a particular example, let

$$\phi_0(x) = \begin{cases} 1, & 0 \leq x < 1 \\ 0, & \text{otherwise} \end{cases} \quad (3.5)$$

When combined with integer translates ($\phi_0(t - k)$ for all $k \in \mathbb{Z}$), this forms a basis set for a function space V_0 . We can denote the span of the set as V_0 ,

$$V_0 = \left\{ \hat{f}(x) : \hat{f}(x) = \sum_{k=-\infty}^{\infty} a[k] \phi_0(x - k) \right\} \text{ for some set of } a[k] \quad (3.6)$$

For the example in equation 3.5, V_0 is the set of all functions that are piecewise constant between integers – which is an approximation of the continuous $f(x)$. That is,

$$\hat{f}(x) = \mathcal{A}_0 \{f(x)\} \quad (3.7)$$

where $\mathcal{A}_0 \{f(x)\}$ is the approximation of $f(x)$ within the space V_0 . If we want to increase the size of the spanned space, and thus improve the approximation, we can scale ϕ_0 to a smaller support. Using a factor of 2 reduction in size of the basis function, the example from equation 3.5 is

$$\phi_1(x) = \begin{cases} 1, & 0 \leq x < \frac{1}{2} \\ 0, & \text{otherwise} \end{cases} \quad (3.8)$$

The space spanned by (half-integer) translates of ϕ_1 will include all of functions that are piecewise constant between half-integers. Obviously this set, V_1 includes V_0 as a subset. We can continue to reduce the size of the ϕ and increasing the space spanned. We can formalize this by defining $\phi_j(x) = 2^{j/2} \phi_0(2^j x)$, where the j is an integer that represents the spatial extent of the basis function – this is also called the scale. Including all necessary translations to cover the real line, indexed by k , define

$$\begin{aligned} \phi_{j,k}(x) &= \phi_j(x - k) \\ &= 2^{j/2} \phi_0(2^j x - k) \end{aligned} \quad (3.9)$$

The $2^{j/2}$ multiplier in equation 3.9 is a normalization factor that will maintain a Parseval's equality, while the $2^j x$ provides the scaling (contraction or expansion of spatial extent) of the original ϕ_0 . Note that shape of the $\phi_{j,k}$'s is invariant - only spatial extent and a normalizing constant multiplier change. Thus ϕ_0 is called the basic or 'mother' scaling function.

If we let $j \rightarrow \infty$, the space spanned by $\{\phi_{j,k}\}$ will approach $L^2(\mathbb{R})$ and the approximation approaches the original signal. For the above example this was initially shown in 1910 by Haar. In practice, since digital signal processing uses a sampled representation (possibly of a continuous function), there is a maximum scale of interest, above which the extra detail is not reflected in the samples. Thus, there is a maximum scale J ($J < \infty$) of interest.

Notice that $V_j \subset V_{j+1}$ for all j , since as the original basis is scaled to higher j (smaller extent, more translations) a more detailed approximation is available, without sacrificing the coarser detail. Thus the spaces V_j form nested function spaces, with each containing all lower indexed function spaces. That is, the space spanned by any $\{\phi_{j,k}\}$ will be a subset of the space spanned by $\{\phi_{j+1,k}\}$, which is the higher scale version of the mother scaling function.

By the nested subset nature, since $\phi_j \in V_j$, we also know that $\phi_j \in V_{j+1}$. However, since $\phi_j \in V_{j+1}$ we must be able to express it as a weighted sum of basis functions for V_{j+1} which are ϕ_{j+1} :

$$\phi_j(x) = \sum_n h_0[n] \phi_{j+1}(x - n) \quad (3.10a)$$

$$= \sum_n h_0[n] \sqrt{2} \phi_j(2x - n) \quad (3.10b)$$

where the h_0 coefficients are the basis weights. The h_0 describe the linkage between the two scales - what weights to apply to a combination of detailed scaling functions to yield the coarser scaling function. A relationship of this type is required for approximations to form nested subspaces. This is a more stringent requirement than

used in the Laplacian Pyramid. These discrete h_0 coefficients will turn out to be a sufficient representation of the continuous scaling function $\phi_0(x)$. As a normalization convention, it is typically assumed that $\sum |h_0|^2 = 1$.

Thus, in order to provide a transform representation of a signal, we need to determine, for a given j , the $a_j[k]$ coefficients in the expansion

$$f(x) \approx \mathcal{A}_j \{f(x)\} = \sum_{k=-\infty}^{\infty} a_j[k] \phi_j(x-k) \quad (3.11)$$

The a_j are directly calculable by a projection:

$$\begin{aligned} a_j[k] &= \langle f(x), \phi_{j,k}(x) \rangle \\ &= 2^{j/2} \int_{-\infty}^{\infty} f(x) \phi_0(2^j x - k) dx \end{aligned} \quad (3.12)$$

However, it is straightforward to develop a simpler method to determine $a_j[k]$ based on the $h_0[n]$ coefficients that define the scaling function. To derive the relationship, set $t = 2^j x - k$ in equation 3.10b,

$$\phi_j(2^j x - k) = \sum_n h_0[n] \sqrt{2} \phi_j(2^{j+1} x - 2k - n) \quad (3.13)$$

changing variables $m = 2k + n$ and substituting this into the inner product (equation 3.12) and re-arranging terms yields

$$a_j[k] = \sum_m h_0[m - 2k] \left[\int f(x) 2^{(j+1)/2} \phi(2^{j+1} x - m) dt \right] \quad (3.14a)$$

$$= \sum_m h_0[m - 2k] a_{j+1}[m] \quad (3.14b)$$

$$= \sum_m h_0[-(2k - m)] a_{j+1}[m] \quad (3.14c)$$

using the definition of a_j from equation 3.12.

That is, the a_{j+1} coefficients along with h_0 can be used in a convolution-type calculation to produce the a_j coefficients (i.e. the higher scale data is used to compute the lower scale data). This convolution varies from the standard convolution due to

the h_0 index which has a negative sign and a multiplier of 2 on k . This implies that the exact relationship is a convolution with $h_0(-n)$ and then the result is downsampled by a factor of 2 – as can be seen by letting $k' = 2k$ and taking only integer indices on $a_j[k']$.

At each stage, N of the a_j coefficients (approximation of $f(x)$ at scale j) are used to calculate $\frac{N}{2}$ of the a_{j-1} coefficients (approximation of $f(x)$ at the coarser scale $j - 1$). This changes the calculation of coefficients from a continuous integral (equation 3.12) to a convolution. The requirement is however, that we know the a_j at the highest detail level, as they are used to generate the coefficients at each coarser level. This will be addressed in section 3.3.4 below.

While this section was motivated by the use of the Haar scaling function, as defined by equation 3.5, the development is valid more generally. This allows great flexibility in choosing the actual scaling function to use in a particular application. For a more complete discussion of the properties of scaling functions and their design see [31, 32]. For the purposes of this work, I will highlight two of the more important facts for scaling functions that fulfill the multiresolution criteria, equation 3.10b.

- The integral of a scaling function will be $\int \phi_0(x)dx = 1$. By equation 3.9, this will also hold for all scales and translations, i.e. $\int \phi_{j,k}(x)dx = 1$, for all j, k .
- $\sum h_0[n] = \sqrt{2}$.

3.3.3 Detail - Wavelet Coefficients

The scaling function allows us to take a signal or image and compute successive coarser scale approximations. What it does not give us is the ability to capture the details between the approximations at two scales. In the Laplacian pyramid, this was captured by a simple subtraction of the coarse approximation from the original. However, due to aliasing this does not yield a good model as to what information

the 'error' images contain. The redundant Laplacian Pyramid representation is also correlated across levels. For the wavelet transform, there is a more precise manner to calculate the multiresolution images.

In order to motivate this, let us look again at the function spaces V_j . We already know this is a set of nested function spaces,

$$\emptyset \subset \cdots V_{j-1} \subset V_j \subset V_{j+1} \cdots \subset L^2(\mathbb{R}) \quad (3.15)$$

But now examine the difference spaces, W_j defined as the orthogonal complement of V_j in V_{j+1} . This is the space spanned by V_{j+1} but not contained in V_j , i.e. what must be added to V_j to get V_{j+1} . Mathematically $W_j \oplus V_j = V_{j+1}$ or $W_j = V_{j+1} \ominus V_j$. The difference space contains the part of the signal present in the more detailed approximation (scale $j+1$), but not in the next coarsest space (scale j). We can recursively decompose the function space V_J as

$$\begin{aligned} V_J &= W_{J-1} \oplus V_{J-1} \\ &= W_{J-1} \oplus (W_{J-2} \oplus V_{J-2}) \\ &\vdots \\ &= \left(\sum_{j=j_0}^{J-1} W_j \right) \oplus V_{j_0} \end{aligned} \quad (3.16)$$

The decomposition is stopped at some coarsest scale of interest, j_0 , defined either by the application or spatial extent of the signal. Thus the function space containing our signal is composed of a sum of $J - j_0$ difference spaces (W_j 's) and a single function space. Since each W_j is contained in V_{j+1} we can expand a set of the basis vectors of W_j in terms of the basis vectors of V_{j+1} , namely the $\phi_{j+1,k}$:

$$\begin{aligned} \psi_j(x) &= \sum_n h_1[n] \phi_{j+1}(x - n) \\ &= \sum_n h_1[n] \sqrt{2} \phi_j(2x - n) \end{aligned} \quad (3.17)$$

where h_1 are the expansion coefficients. ψ_j is also called the wavelet function, and $\{\psi_{j,k}\}$ will form a basis for the difference space W_j . ψ_0 , a weighted sum of the mother scaling function, is called the mother wavelet function. Similar to scaling functions, wavelets at any scale are invariant in shape, varying only in magnitude and spatial extent. Note if ϕ has finite extent, ψ will also have finite extent. Infinite extent scaling functions and wavelets (that converge to zero as we move towards infinity) are possible but are not discussed here.

Assuming that $f(t) \in V_J$, from equation 3.16 we can expand it in terms of the basis functions of the W_j 's and a V_{j0} . Using the fact that the ψ_j 's span W_j and ϕ_j 's spans V_j , we develop the wavelet transformation equation from a weighted sum of these basis functions, including the necessary translations in k :

$$V_J = \left(\sum_{j=j_0}^{J-1} W_J \right) \oplus V_{j_0} \quad (3.18a)$$

$$f(x) = \sum_k \sum_{j=j_0}^{J-1} d_j[k] \psi_{j,k}(x) + \sum_k a_{j_0}[k] \phi_{j_0,k}(x) \quad (3.18b)$$

Using an approach similar to above for the approximation coefficients, we can determine the wavelet transform coefficients, $d_j[k]$ by a projection:

$$d_j[k] = \langle f(x), \psi_{j,k}(x) \rangle = 2^{j/2} \int_{-\infty}^{\infty} f(x) \psi(2^j x - k) dx \quad (3.19)$$

which again we can simplify through similar algebra to

$$d_j[k] = \sum_m h_1(-(2k - m)) a_{j+1}[m] \quad (3.20)$$

Thus the wavelet coefficients at any scale are calculable from the scaling coefficients at the next finer scale by a convolution relationship with h_1 similar to equation 3.14c.

The a_j coefficients are termed the approximation or scaling coefficients as they are the approximation of the signal in the V_j function space. The d_j coefficients are termed the detail or wavelet coefficients as they contain the detail of the signal within

W_j . As compared to the Laplacian pyramid, the wavelet coefficients provide better decorrelation of subbands and eliminate the redundancy.

The wavelet is completely defined by the scaling function as shown in equation 3.17. Using this and the properties of the scaling coefficients we can derive the following properties:

- The integral of a wavelet function will be $\int \psi_0(x)dx = 0$. This will also hold for all scales and translations, i.e. $\int \psi_{j,k}(x)dx = 0$, for all j, k . Thus the $d_j[k]$ are zero-mean.
- For an orthogonal scaling function, $h_1[n] = (-1)^n h_0[1-n]$. Thus h_0 is a sufficient representation of both ϕ and ψ .
- A Parseval's relationship holds: $\int |f(t)|^2 = \sum_k \sum_{j=j_0}^{\infty} |d_j[k]|^2 + \sum_k |a_{j_0}[k]|^2$

The modified convolution with h_0 and h_1 leads to a simpler calculation of expansion coefficients than a continuous inner product. However, it is interesting to look at various scaling/wavelet function pairs, especially if we remember that the DWT calculates the inner product of the data with these functions. Figure 3.5 shows the simplest scaling/wavelet function which was defined above in equation 3.5. This is called the Haar basis. Figure 3.6 shows two of the Daubechies wavelet/scaling function pairs. These were originally defined by Daubechies, and are designed to be the shortest spatial extent functions which meet the necessary criteria with a given number of vanishing moments. Parts (a) and (c) show the Daubechies - 4 functions, with a zero first moment. Parts (b) and (d) show the Daubechies - 12 functions with 5 vanishing moments. There will be $\frac{N}{2} - 1$ vanishing moments for a Daubechies wavelet of length N . The importance of vanishing moments relates to the ability to sparsely represent a continuous polynomial function. The Haar wavelet is actually the Daubechies -2 wavelet, although it maintains its own name. Note that

the lower Daubechies wavelets are not smooth (The Haar wavelet is even discontinuous). Finally, figure 3.7 shows another pair of wavelets/scaling functions that form a biorthogonal pair. This concept will be discussed later, but notice these are symmetric, as opposed to all other orthogonal pairs, with the exception of the Haar wavelet. The h_0 coefficients are given in Table 3.1.

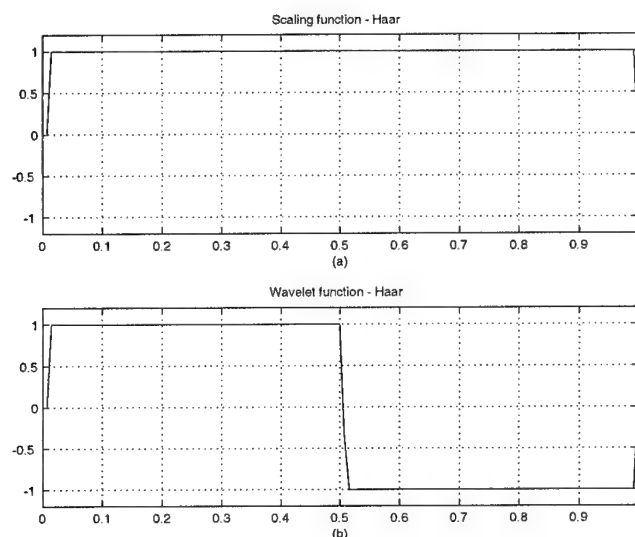


FIGURE 3.5. The Haar basis: (a) scaling function and (b) wavelet function

Wavelet Name	h_0 coefficients
Haar	0.707, 0.707
Daubechies - 4	0.483, 0.836, 0.224, -0.129
Daubechies - 12	0.111, 0.495, 0.751, 0.315, -0.226, -0.129, . . . 0.097, 0.027, -0.032, 0.001, 0.005, -0.001
Biorthogonal(1)	0.038, -0.024, -0.111, 0.377, 0.853, 0.377, -0.111, -0.024, 0.037
Biorthogonal(2)	-0.065, -0.041, 0.418, 0.788, 0.418, -0.041, -0.065

TABLE 3.1. Filter coefficients for various wavelet bases

3.3.4 The Discrete Wavelet Transform

The algorithm above is based on knowing what the approximation coefficients are at the finest scale of interest. Since computation typically involves discrete sampled

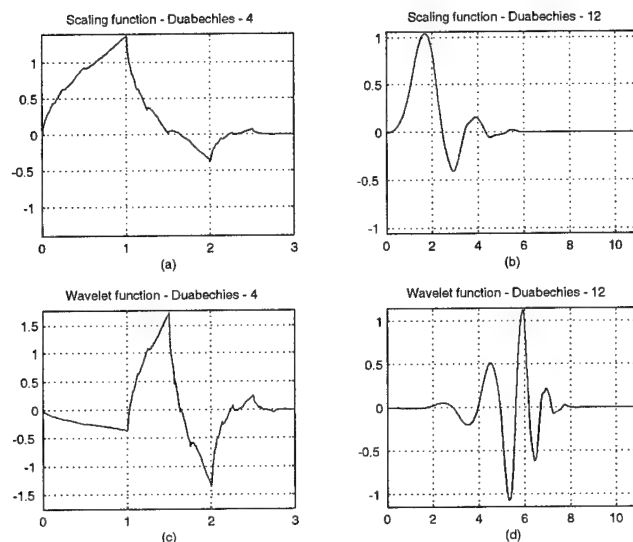


FIGURE 3.6. The Daubechies-4 and Daubechies-12 basis functions. (a) and (c) are for Daubechies-4, (b) and (d) are for Daubechies-12

data, an understanding of the linkage between the original continuous signal, the discrete sampled signal, and the approximation coefficients is important. This finest scale of interest is typically based on the sampling of the original signal. One obvious choice is to use the signal samples as the approximation coefficients, which turns out to be a very good approximation for most signals.

From a continuous signal, the exact scaling coefficients at scale j are found by projecting the signal onto V_j , i.e. $a_j[k] = \langle f(x), \phi_{j,k}(x) \rangle$. If $f \in V_j$ this will be an exact representation of $f(x)$. However, this can not be guaranteed for an arbitrary signal. However, if we take the signal samples, $f[n] = f(nT)$ for signals sampled at or above the Nyquist rate, these are a good approximation to the scaling coefficients $a_j[k]$. In [33], it is shown that such samples are a third order approximation to the exact scaling coefficients for weak constraints of the wavelet system used, and increasing the sampling rate by a factor of M , will reduce the error in the coefficients by a factor of M^3 . This can be motivated intuitively by noting that, for fine enough scale, the scaling function can be approximated by a Dirac impulse at its center, since

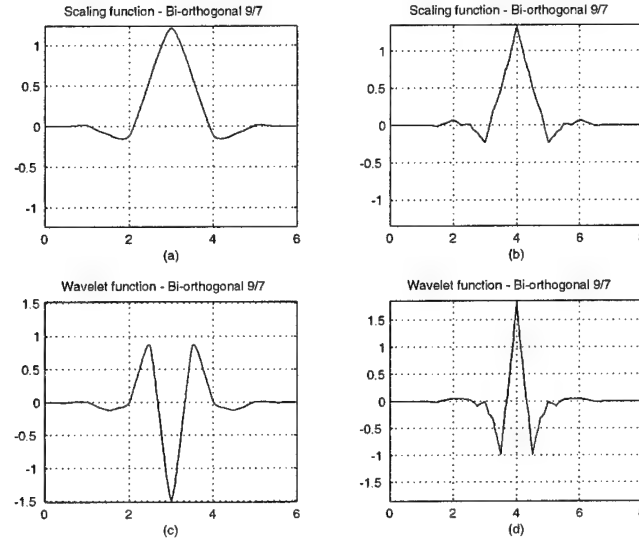


FIGURE 3.7. The bi-orthogonal 9/7 basis functions. (a) and (c) form a pair; (b) and (d) are the other pair.

$\int \phi_j(x)dx = 1$ and as the scale increases, the support is shrinking. Knowing that a perfect sampling system is a projection onto a Dirac impulse, as the scale increases, we approach the limit. Mallat[31, p. 257] discusses how this can be extended to account for finite resolution sampling (i.e. pixels) when calculating the scaling coefficients. For the current work, I will use the pixel values as an adequate representation of the scaling coefficients at the finest scale.

From equations 3.14c and 3.20, the algorithm to calculate the discrete wavelet transform (DWT) is

1. Define $a_J[k]$ as the pixel values.
2. $a_{J-1} = \downarrow (a_J \otimes h_0[-n])$, where \downarrow is downsampling by a factor of 2 (removal of all even samples).
3. $d_{J-1} = \downarrow (a_J \otimes h_1[-n])$.
4. If it is desired to transform to the next level, repeat the above procedure starting

step 1 with only the $a_{J-1}[k]$.

At each stage, N approximation coefficients are transformed to $\frac{N}{2}$ wavelet coefficients ($d_j[k]$) and $\frac{N}{2}$ scaling coefficients ($a_j[k]$) – which is the same number as the input.

The inverse DWT (IDWT) is calculated by the following algorithm

1. Start with $a_J[k]$ and $d_J[k]$.
2. $a_{J+1} = (\uparrow a_J) \otimes h_0$, \uparrow is upsampling by 2 (insertion of zero after each sample).
3. $d_{J+1} = (\uparrow a_J) \otimes h_1$.
4. $a_{J+1} = a_{J+1} + d_{J+1}$.
5. Repeat the above procedure starting step 1 with a_{J+1} and d_{J+1} until original scale is reached.

Note that the same filters are used for the DWT and the IDWT, with the exception of flipping the filters for the DWT, i.e. using $h_0[-n]$ rather than $h_0[n]$. Also, the borders must be addressed for finite length signals. This can be dealt with three ways:

1. Assuming periodic extension of the a_j vector, which is identical to circular convolution. However, the discontinuity at the edge will produce large wavelet coefficients that relate to the periodic extension, not the signal itself.
2. Assuming a folded symmetric extension of the signal at the edges. This approach reduces, but does not necessarily eliminate the issue of larger wavelet coefficient magnitudes near the borders.
3. Defining 'boundary wavelets'[31] in addition to the original wavelets, which are used near the border to maintain vanishing moments, and thus minimal wavelet magnitude, at the boundary.

This choice is critical in applications such as compression where large coefficients will be coded. If they reflect edge discontinuity rather than signal content, this will lead to wasted bits. However, in signal and image processing, the impact will largely be restricted to the edges and thus not a large problem. In the following work, I'll use the first approach except where specified when I'll use the second.

The forward and inverse DWT algorithms are shown graphically in figure 3.8.

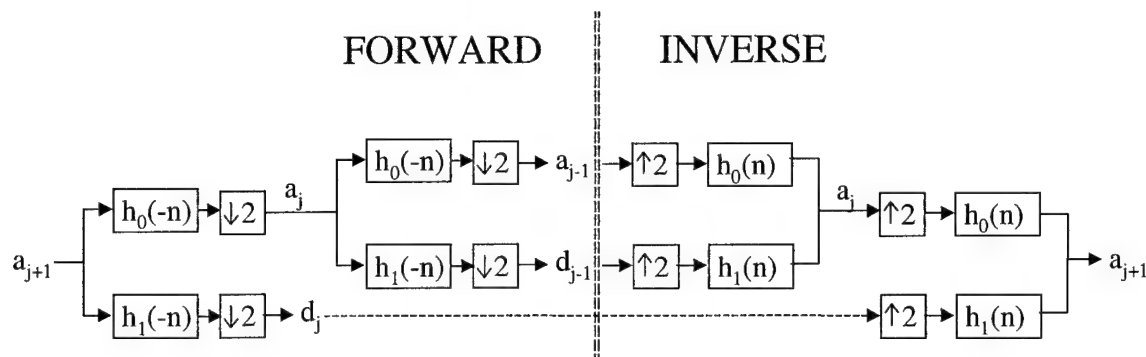


FIGURE 3.8. Algorithm for forward and inverse discrete wavelet transform

The 2-D DWT uses a separable implementation of the 1-D filters along the rows and columns. However, this will yield four bands, rather than the two bands (a_j 's and d_j 's) for a 1-D signal since we have four combinations of the filters applied to the row/column. Skipping ahead, we will see that the scaling coefficients are a low-pass representation and the wavelet is a high-pass. Thus each of the four subbands from the 2-D image will be represented by two letters, LL, HL, LH, HH, where the first letter specifies whether a low- or high-pass filter is used on the rows, and the second letter represents the columns. The subbands are usually displayed as in figure 3.9(a) where a two-scale transform has been shown. Note that only the LL band is iterated upon at each subsequent transform level and is replaced by the 4 decomposed subbands. An example of a 2-D DWT is shown in figure 3.9(b) for the Lenna image. Note that the magnitude of the wavelet coefficients is shown to improve visibility.

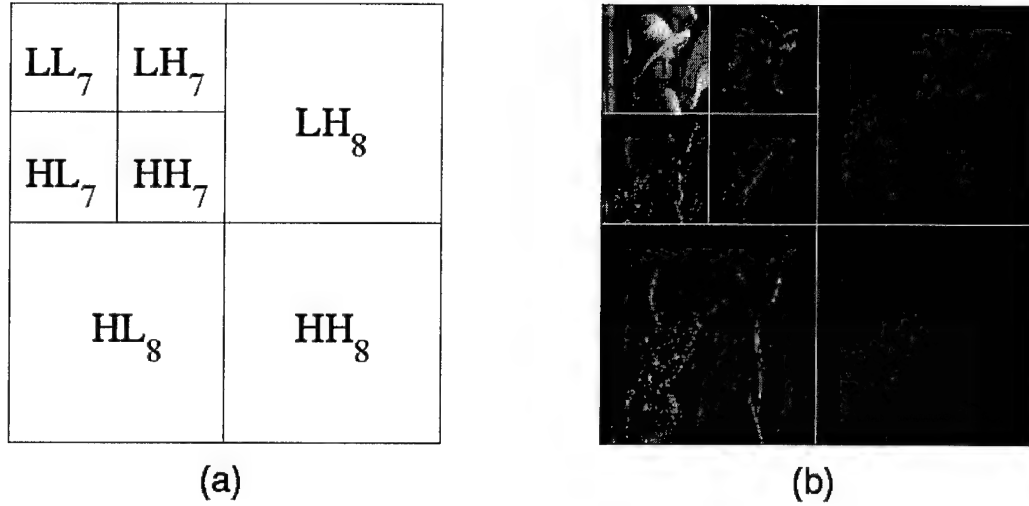


FIGURE 3.9. 2-D DWT display: (a) is location of subbands in display. (b) is DWT of Lenna image

3.4 Wavelet discussion - Frequency Domain

The above derivation has been motivated and derived strictly from a spatial domain point of view. However, as alluded to before, there is a strong connection between multiresolution techniques and frequency domain concepts. High frequency content would correspond to a high scale, whereas lower frequency corresponds to a lower, coarser scale. In this section I'll develop this relationship in more detail. Both perspectives are important for a full understanding of the DWT.

The Fourier Transform is a well-known signal decomposition. In the discrete case, the Discrete Fourier Transform (DFT) is calculated through projection onto complex sinusoids,

$$F[k] = \sum_{n=-\infty}^{\infty} f[n] e^{-i \frac{2\pi k}{N} n} \quad (3.21)$$

Each of the $F[k]$ terms represents the frequency content in f , corresponding to the digital frequency around $\omega = \frac{2\pi k}{N}$. One aspect of this representation is that there is no time localization of the frequency content and vice versa. Thus, a plot of the DFT

magnitude will give information as to which frequencies are in the signal, but lacks any information as to the location and duration in time. As an example, suppose we have two signals, each with 25 Hz and 50 Hz tones. The first has 5 sec of 25 Hz tone and the 5 sec of 50 Hz tone; the second is 10 sec of superposition of 25 and 50 Hz tone. The DFT magnitude does not provide a concise difference between the cases - both will have energy in the 25 and 50 Hz bins. While the phase does provide the necessary information to invert the transform and recover the exact time domain representation, it does not give any time localization of frequency content. Thus, DFTs are best suited for stationary signals where the frequency content remains constant in time.

This situation is graphically demonstrated below in figure 3.10 which is a time-frequency distribution graph. A 1-D signal is represented in a 2-D plot, along the time and frequency axes. Each rectangle represents a particular sample. In the time domain (figure 3.10a), we know exactly where the energy is located (sequentially by sample), but no localization at all in the frequency domain - the energy content in any frequency band is indeterminate. The opposite effect occurs for the Fourier coefficients (figure 3.10b): we know exactly what frequency band the energy corresponds to, but no information of temporal location is known.

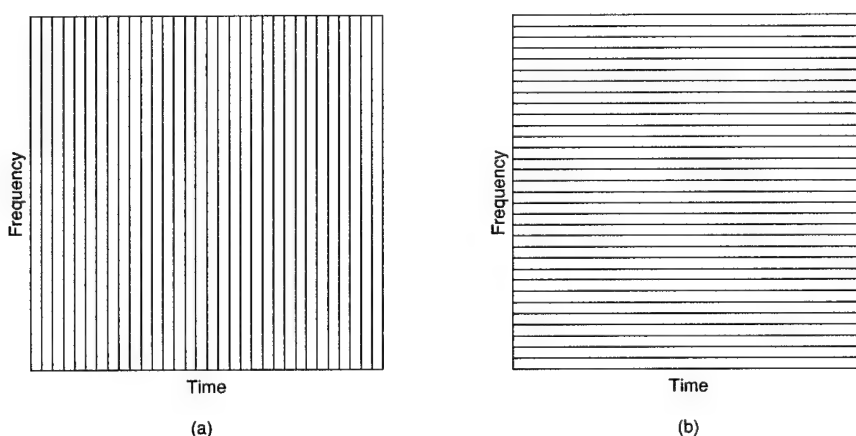


FIGURE 3.10. Time-frequency distribution for a 1-D signal in the (a) time domain and (b) Fourier domain

One approach to mitigate this problem is to use Short-Time Fourier Transforms (STFTs). The concept is to window the function within a finite interval, take the DFT, and then move the window (the translated window typically overlaps part of previous location). By doing this, we end up with DFT coefficients (i.e. frequency content) at each translation. In terms of time-frequency plots, frequency content is somewhat localized. Graphically this is shown in a time-frequency plot in figure 3.11. The amount of localization possible is governed by a fundamental Heisenberg-type limit which states $\Delta t \Delta \omega$ must always be greater than a constant. Thus, we can never exactly locate a frequency in time, and the more we try, the worse the frequency resolution becomes. Although the blocks are shown with solid borders, this is simply the standard deviation of the extent in time and frequency. One additional issue with STFT's is that the windowing functions can introduce artifacts in the transform domain.

One aspect to note about the STFT is that each of time-frequency 'atoms' will be of constant size. However, in many applications, the high frequency content (e.g. an edge) is short in duration while the low frequency content (e.g. drifts in background) changes slowly. Thus we often desire better localization of high frequencies and can live with poorer spatial/time resolution for low frequencies.

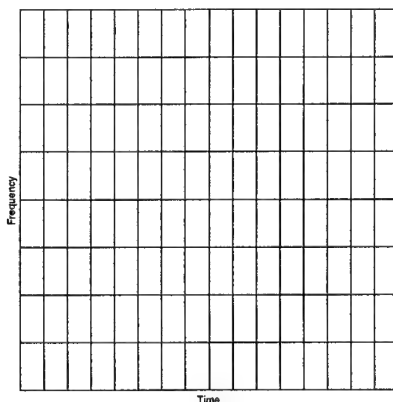


FIGURE 3.11. Time-frequency distribution for 1-D signal using a STFT

Another time-frequency analysis tool is the Wigner-Ville distribution. This leads to a 2-D plot of the energy density at a position u , and frequency ξ :

$$W_f(u, \xi) = \int_{-\infty}^{\infty} f\left(u + \frac{\tau}{2}\right) f^*\left(u - \frac{\tau}{2}\right) e^{-i\tau\xi} d\tau \quad (3.22)$$

While useful in some applications, it suffers from the fact that it is a quadratic term which leads to interference terms that are not related to the signal, but rather an artifact of the distribution.

The DWT provides another alternative approach. In the DWT, the coefficients are calculated via a convolution type relationship as discussed in the previous section. Since convolution implies multiplication in the Fourier domain, let us look at the frequency response of the ‘filters’ h_0 and h_1 which correspond to ϕ and ψ . From the concept of approximation, we can expect h_0 to be a low-pass filter. While not providing a formal proof, this can be seen since $\sum h_0(n) = \sqrt{2}$, $H_0(0) = \sqrt{2}$. The definition of ϕ also requires that $H_0(\pi) = 0$. For an orthogonal basis, $h_1(n) = (-1)^n h_0(1 - n)$. Thus, h_1 will be related to h_0 as the frequency shifted version, $H_1(0) = 0$ and $H_1(\pi) = \sqrt{2}$. This implies that h_1 is a high-pass counterpart. A plot of the frequency response magnitude for the Daubechies - 12 wavelet is in figure 3.12.

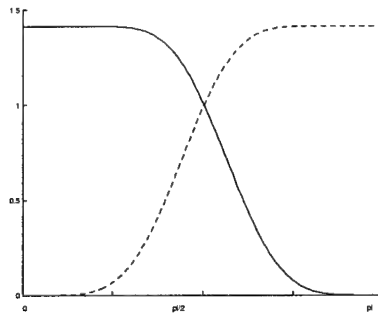


FIGURE 3.12. Frequency response of Daubechies-12 filters. The scaling coefficient filter is solid, the wavelet coefficient filter is dashed.

Since we are downsampling after each of the filters, a natural question is what effect does aliasing have. If we calculate a DWT and then the inverse DWT, accounting

for the effects of downsampling and upsampling, the aliasing will cancel if

$$|H_0(\omega)|^2 + |H_0(\omega + \pi)|^2 = 2 \quad (3.23)$$

This also implies that for FIR filters,

$$h_1(n) = (-1)^n h_0(1 - n) \quad (3.24)$$

This interesting result shows that enforcing cancellation of aliasing is consistent with the derivation using spatial domain multiresolution function spaces.

A look at the frequency response of the DWT leads to an understanding of how the DWT tiles the time-frequency plane. The first level of wavelet coefficients, due to the high-pass nature, will be the information from the top half of the frequency domain, $\frac{\pi}{2} < \omega < \pi$. Note that as the filters are FIR, this is not a sharp cut-off at $\frac{\pi}{2}$, but for the discussion here it will serve as a reasonable approximation. The next scale will be a band-pass, $\frac{\pi}{4} < \omega < \frac{\pi}{2}$. This can be understood from the downsampling of the approximation coefficients (a_j 's). Due to downsampling, the digital frequency of π for the d_{J-1} coefficients relates to $\omega = \frac{\pi}{2}$ in the original signal (d_J coefficients). Thus, a high-pass filter on the downsampled data extracts the frequencies $\frac{\pi}{4} < \omega < \frac{\pi}{2}$ of the original signal. Subsequent downsampling will shift the band-pass in a similar manner. Note again, that the band-pass will not be a sharp cut-off.

The time-frequency plot of the DWT coefficients is shown in figure 3.13, where we see the multiple samples related to the higher scale frequencies and fewer coefficients related to the lower frequencies, in a dyadic (i.e. factor of two) relationship, as expected due to downsampling. Note that frequency axis has been labeled scale to point out the lack of sharp cut-off in the filters and thus the data. The exact relationship between frequency and scale can be derived by the frequency response of the filter h_1 . The highest scale coefficients (top) are the $d_J[k]$ terms, the next band down are the $d_{J-1}[k]$ terms, and so on. For a signal of length N , we find that $J = \log_2 N$ and there are $\frac{N}{2} d_J[k]$ terms, $\frac{N}{4} d_{J-1}[k]$ terms and so on. For a complete

decomposition of J levels, the approximation output will be a single value. Since all $d_j[k]$ are zero-mean, the remaining $a_j[k]$ term(s) are the DC term – thus the name scaling coefficient as it scales the overall magnitude of the signal.

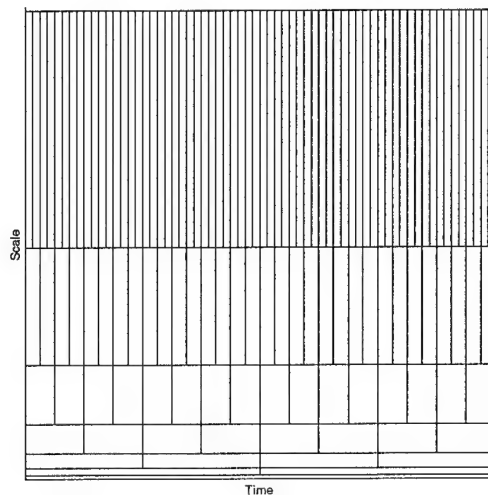


FIGURE 3.13. Time-frequency distribution for a 1-D signal in the wavelet domain

In order to facilitate discussions, the following terms are commonly used and can be visualized in figure 3.13. The *children* of a wavelet coefficient are the higher scale wavelet coefficients that share the same temporal location – every coefficient (except the highest scale) will have two children. The *parent* of a wavelet coefficient is the coefficient at the next lower scale which shares the same temporal location – every coefficient (except the coarsest scale) will have one parent. The *ancestors* of a coefficient are all of the lower scale coefficients that share the same temporal location (parent, parent’s parent, etc.). The *descendents* are all of the shared higher scale coefficients (children, children’s children, etc.).

Chapter 4 will discuss particular applications of the DWT. However, there are certain properties of the DWT that bear discussion here. These qualities of the DWT lead to their success in the applications discussed below.

P1: Locality: As discussed above, wavelet coefficients represent a localized content

in both time and frequency. Thus modification of a wavelet coefficient will have only a local effect on the signal. Alternatively, modification in the Fourier domain has a global effect.

P2: Multiresolution: The transform coefficients provide insight into signal properties across a set of scales.

P3: Sparse Representation: DWTs tend to be sparse and have few large coefficients that denote signal information at that scale.

P4: Decorrelation: The DWT approximately decorrelates the signal.

From these properties and using the self-similarity of natural images, the following additional properties are known.

Distribution: DWT coefficients are distributed as a large peak near zero with heavy tails. They are not Gaussian in distribution, but either a Generalized Gaussian distribution (GGD) or a Gaussian Mixture model (GMM) can closely approximate them.

Magnitude decay: Coefficient magnitudes decay exponentially across scales.

Persistence/clustering: Large/small magnitudes tend to propagate across scales and be clustered within the same scale.

3.5 Variations on a theme

The DWT discussed above is the most straightforward implementation. Multiple variations of the ideas discussed above have been developed to extend the concepts for various reasons. This section will discuss a few of them that are critical to my research.

3.5.1 Biorthogonal Wavelet Transforms

Biorthogonal wavelet transforms are generalization of orthogonal transforms described above. One of the keys is that, similar to a biorthogonal basis above, the decomposition filters h_0 and h_1 are not the same as the synthesis filters, denoted as g_0 and g_1 . The modification to the orthogonal algorithm of figure 3.8 is shown in figure 3.14. To ensure perfect reconstruction, we arrive at a similar relationship to equation 3.24,

$$h_0[n] = (-1)^n g_1[1-n] \quad (3.25a)$$

$$h_1[n] = (-1)^n g_0[1-n] \quad (3.25b)$$

This shows that the low-pass decomposition filter is conjugate to the high-pass synthesis filter.

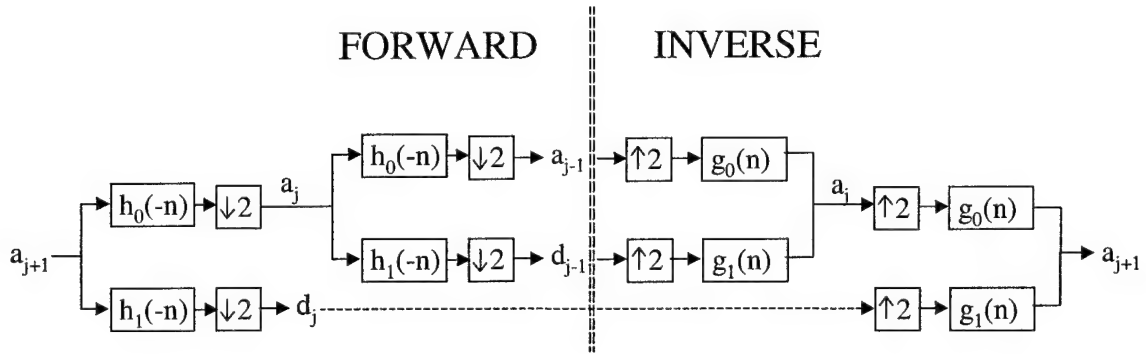


FIGURE 3.14. Algorithm for forward and inverse discrete wavelet transform using bi-orthogonal wavelets

The benefit of this type of transform is that there are more degrees of freedom in designing biorthogonal wavelet bases than in the orthogonal case due to the addition of the g filters. One of the primary results of this is that symmetric biorthogonal wavelets are possible. In the orthogonal case, the Haar wavelet is the only possible symmetric basis.

There are drawbacks to the biorthogonal transform. First, there is no Parseval type equality. Second, while an orthogonal transform of Gaussian noise remain Gaussian, this is not true for biorthogonal transforms. Some biorthogonal constructions seek to minimize these effects by being ‘almost’ orthogonal (A close to B in equation 3.2), but there is no way to eliminate them. The 9/7 Biorthogonal wavelet set (the one shown in figure 3.7 above) is often used as a standard choice in compression algorithms as it works well for a large class of images. The frequency response of the biorthogonal 9/7 filter set is shown in figure 3.15.

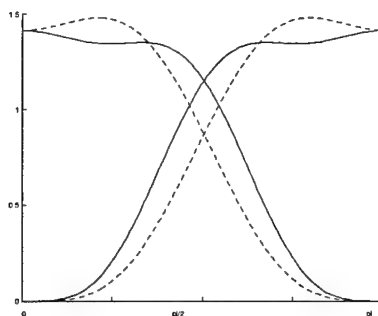


FIGURE 3.15. Frequency response of bi-orthogonal 9/7 filters. The first pair of low-pass/high-pass filters is solid, the second pair is dashed.

3.5.2 Wavelet Packets

The above DWT's are based on decomposition of only the low-pass output. However, we can also decompose the high-pass output (d_j 's) as well. If we do this at each branch, we end up with a binary tree for the algorithm. The time-frequency tiling for this would then be similar to the STFT, where all of the atoms are of a uniform size. The effect of this is to further separate each of the detail vector spaces, W_j . This allows great freedom in parsing out the time-frequency plane. For certain applications, this will be necessary and/or beneficial to account for particular system or signal characteristics.

3.5.3 Undecimated Wavelet Transforms

The DWT as discussed above is a linear transform, but is not shift invariant due to the down sampling. This is problematic for many applications such as computer vision, where a simple shift in the image should not change the transform data other than a similar shift. A straightforward way to achieve this is to remove the downsampling and upsampling steps - termed an *undecimated DWT*. This is also called the *redundant* transform or a *stationary* transform. However, this also requires changes to the filters to account for the lack of down/upsampling. In the spatial domain, the downsampling implies that the filter will be applied only to every other sample at the next level of decomposition. To account for this, the filter must be upsampled by a factor of two at each level by zero insertion. This is called the *algorithm à trous* or ‘algorithm with holes’ as a reference to the sparsity of the filter as the decomposition progresses. Another method is to calculate the decimated DWT for each shift of the signal and interleave the results. At the first level this requires two DWT’s due to the elimination of half the samples at each downsampling. At the second level, it will require 4 shifts, two for each of the two first level shifts. Shensa[34] demonstrates the equality of the algorithm à trous and interleaving decimated DWT’s.

The undecimated DWT provides a shift invariant transform, but at the expense of redundancy. For a N -length signal with M levels of decomposition, $(M + 1) * N$ samples will be output. As will be demonstrated below, this can markedly improve the robustness of algorithms, especially denoising.

The *algorithm à trous* also provides a third perspective on the DWT parsing of the time-frequency plane shown in figure 3.13. Upsampling the filters changes their frequency response by contracting the frequency axis by a factor proportional to the upsampling factor. The first wavelet decomposition is calculated by high-pass filtering the data, therefore the predominant frequencies at this level will be $\frac{\pi}{2} < \omega \leq \pi$. Upsampling h_1 will change the frequency response by contracting the

frequency axis by a factor of 2. That is, the peak at $\omega = \pi$ will be at $\omega = \frac{\pi}{2}$. The assumed periodic nature of the DFT will then imply the filter will fall off to zero at $\omega = \pi$ (dashed line in figure 3.16(a)). At the next (third) level of decomposition, the peak will be at $\frac{\pi}{4}$ and the periodic extension implies a new peak at $\omega = \frac{3\pi}{4}$ as seen in the dotted plot in figure 3.16(a). However, since the first and second decompositions removed most of the energy in frequencies above $\frac{\pi}{2}$, the wavelet coefficients at the third decomposition level have little energy above $\omega = \frac{\pi}{2}$. Figure 3.16(b) shows the resultant frequency content of a transform for an input of constant frequency magnitude, which demonstrates this fact. For the plots in figure 3.16, a Daubechies-20 wavelet system was used.

Generalizing this, for a signal with maximum scale J , the primary frequency content at scale j will be

$$\frac{\pi}{2^{J-j}} < \omega \leq \frac{\pi}{2^{J-j-1}} \quad (3.26)$$

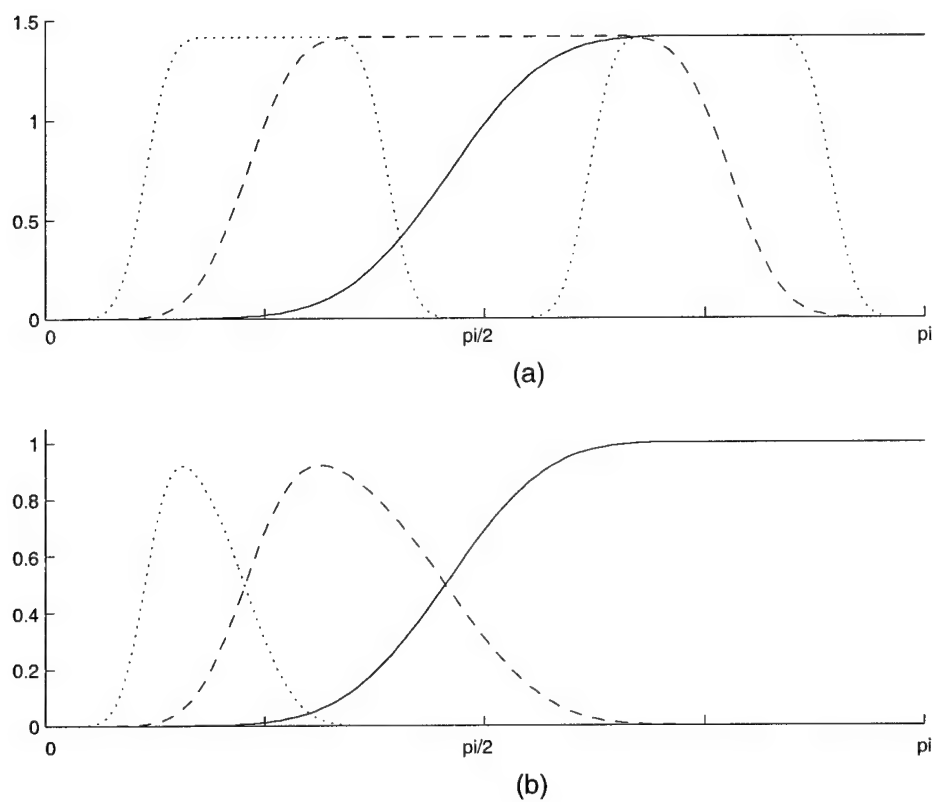


FIGURE 3.16. (a) Frequency response of wavelet filters using the *à trous* algorithm. Solid lines are the first level, dashed is second level and dotted is third level in the decomposition. (b) is the resultant band-pass nature of the respective levels for an input with flat frequency content.

CHAPTER 4

IMAGE PROCESSING WITH WAVELETS

4.1 Overview

An understanding of the properties of wavelet transforms, as discussed in the previous chapter, lead to many applications in signal/image processing. In this chapter I will discuss some of these applications as they apply to my work. In general, most algorithms are based on the outline of figure 4.1. The DWT is used to transform the data into a form that allows either a simpler or better (hopefully both) representation of the data in terms of the desired application. After processing the transform data, it is inverse transformed to provide the final result.



FIGURE 4.1. Algorithm for wavelet domain processing

The first section will discuss denoising a signal or image via wavelets. This is necessary background for the next section which will overview image restoration using wavelets. The last section will discuss techniques for image interpolation with wavelets.

4.2 Wavelet Denoising Algorithms

Let us use the model

$$g = f + n \tag{4.1}$$

where g is the observed data, f is the object, and n is the noise. The goal is to determine f given g and some knowledge of the noise.

One solution to this problem is to determine the minimum MSE estimate of f , which is the Wiener estimate. In the Fourier domain it is

$$\hat{F} = WG = \frac{1}{1 + \frac{S_n(\xi, \eta)}{S_f(\xi, \eta)}} G \quad (4.2)$$

where S_n and S_f are the PSD of the noise and signal, respectively. This is the same as equation 2.21b used in image restoration, without the inverse of the blurring operator.

Wavelet domain techniques have produced better results both in terms of lower MSE and subjective visual quality than possible by Wiener estimates. The intuition as to why wavelets provide a good transform to denoise is that the wavelet domain yields a sparse (or compact) representation. This implies that the energy of the signal is in few large coefficients. Since the DWT is a linear transform in an orthonormal basis, AWGN will remain AWGN. Thus the wavelet domain signal energy is combined in few larger coefficients with the same level of noise – making it easier to identify and retain. An interesting side note is that the compactness in representation also leads to good data compression performance. Transform domain denoising and compression share similarities: denoising seeks to identify and retain those coefficients related to the signal and not noise, while compression seeks to code only the most important coefficients, which are those that contain the most information about the signal. In fact many wavelet compression techniques do suppress AWGN noise in the process.

The first approach to wavelet domain denoising was based on thresholding of the wavelet coefficients. The Wiener filter can be looked at as a variable thresholding of the Fourier coefficients – the magnitudes of the frequency components are reduced in relation to the expected SNR. The original wavelet based denoising technique is similar: threshold the magnitude of the wavelet coefficients based on SNR. However, in the wavelet domain, this is accomplished by a simple thresholding – setting coef-

ficients below a threshold to zero. Note that the thresholding is applied to wavelet coefficients only, and not the scaling coefficients. Originally proposed by Donoho and Johnstone[35], thresholding techniques have been shown to be very powerful. Fourier denoising techniques seek to separate the noise from the signal as much as possible, and then attenuate those frequencies based on the SNR prediction from the PSDs. Wavelet techniques are different in that the signal and noise are separated by amplitude and thus are not reliant on a priori signal PSD predictions.

The two approaches to thresholding are

- Hard thresholding: set to zero all coefficients with magnitudes below the threshold without adjusting any others,

$$d_{j,k} = \begin{cases} d_{j,k}, & |d_{j,k}| > T \\ 0, & |d_{j,k}| \leq T \end{cases} \quad (4.3)$$

- Soft thresholding: set to zero all coefficients with magnitudes below the threshold, but also reduce the magnitude of those above the threshold by the amount of the threshold, i.e. shrink all coefficients toward zero by the amount of the threshold.

$$d_{j,k} = \begin{cases} d_{j,k} - T, & d_{j,k} > T \\ d_{j,k} + T, & d_{j,k} < -T \\ 0, & |d_{j,k}| \leq T \end{cases} \quad (4.4)$$

The results will have slightly different characteristics. Hard thresholding will provide better MSE, but may lead to spurious oscillations within the signal. This is due to an induced non-continuity of the wavelet coefficients near the threshold. Soft thresholding retains the original smoothness of the image and contains fewer spurious oscillations (better visual quality).

The calculation of the best threshold is not obvious and involves a similar trade-off as discussed previously with LSI filters. For a low threshold, the result will look similar to the signal, but will retain most of the noise. As the threshold increases, the noise will be reduced, but the resultant signal will become smoothed, due to

removal of some signal energy from the transform domain. As expected, however, the threshold will always be a function of the noise variance.

The ‘universal’ threshold for 1-D ‘smooth’ signals is $T = \sigma\sqrt{2\log_e N}$, with σ the noise variance and N the number of samples[35]. This is derived either from 1) minimax estimation optimization[31] or 2) determining what the highest expected noise amplitude will be. The use of N in the threshold is a result of an assumption in the above derivations that N samples are taken from a continuous function on $[0, 1]$. As the number of samples grows, redundancy in the samples increases as does the probability of large noise values from the tails in the Gaussian distribution. This implies that a higher threshold should be used. However, the universal threshold does not yield the best results either in terms of MSE or visual appearance – it is overly ambitious in noise removal at the expense of signal smoothing. However, it is optimal in the minimax estimation sense, i.e. minimizing the maximum error for any allowable signal.

More generally, for an arbitrary signal, the threshold is calculated as a multiple of the noise standard deviation. For certain classes of signals, this can be optimized in the MSE sense. From experiments on multiple typical signals, the MSE optimal threshold for 1-D signals is usually $1 - 1.5\sigma$, whereas for 2-D signals it is $2 - 3\sigma$. For cases where the noise standard deviation is not known, it can be estimated from the median of the absolute value of the samples[31, p. 459],

$$\sigma = \frac{\text{median}}{0.6745} \quad (4.5)$$

Equation 4.5 is derived from samples of AWGN only. However, the estimate is fairly robust in the presence of some signal as the median is relatively insensitive to the addition of a few large samples.

The above techniques can be further improved by using the UDWT and applying the threshold to it. This will provide improved results, as originally demonstrated by [36]. If we look at the decimated transform, shifting the original signal by one

sample gives a different, but equivalent, representation of the data in the wavelet domain. By using the average of all possible shifts (equivalent to denoising the undecimated transform), the results are improved. This should be expected in the same way multiple noisy measurements have more information than one measurement, assuming constant noise variance.

Simple coefficient thresholding provides a very computationally efficient manner to denoise a signal. However, it ignores additional available information that may be used to increase performance based on statistical models of the data and transform. The next step is based on modifying the thresholds based on models of scale dependencies. In [37, 38] a spatially varying threshold is developed based on the local statistics of the coefficients, motivated by similar work in the compression community. They develop the subband dependent threshold of $T_{j,k} = \sigma^2 / \sigma_{j,k}^2$, where σ^2 is the noise variance and $\sigma_{j,k}^2$ is the wavelet coefficient variance, adapted to each individual coefficient based on the statistics of ‘similar’ nearby coefficients. Changing the threshold based on the wavelet coefficient variance is intuitively reasonable. As the wavelet variance increases, this implies signal information such as an edge – the threshold is decreased to include any signal present, at the possible cost of some noise. In smooth areas (low signal content) with low wavelet variance, the threshold increases to reduce noise.

Ghael and Choi [39, 40] developed the idea of Wiener filtering the wavelet coefficients. They also show how hard thresholding is equivalent to a Wiener filter for the model of the signal estimate given by

$$\hat{d}_{j,k} = \begin{cases} \infty, & |d_{j,k}| > T \\ 0, & |d_{j,k}| \leq T \end{cases} \quad (4.6)$$

The Wiener filter will be optimal in an MSE sense for orthogonal transforms in AWGN, but require an accurate model of the signal.

The next major change was to apply a Bayesian approach of using a probability model of the coefficients[41]. The sharp peak at zero and heavy tails experimentally

observed for natural images rule out Gaussian statistics. The two competing models that show validity are the generalized Gaussian distribution (GGD) and Gaussian mixture model (GMM).

The GGD model is given by:

$$p_{\nu,\sigma}(x) = \frac{\nu\eta(\nu)}{2\Gamma(1/\nu)} \frac{1}{\sigma} \exp\left(-\left[\eta(\nu)\frac{|x|}{\sigma}\right]^\nu\right) \quad (4.7)$$

$$\text{where } \eta(\nu) = \sqrt{\frac{\Gamma(3/\nu)}{\Gamma(1/\nu)}}$$

where σ is the standard deviation, and ν is the shape parameter. Two special cases are $\nu = 2$, the Gaussian distribution and $\nu = 1$, the Laplacian distribution. Written more compactly to see the relation to the Gaussian case,

$$p_{\nu,\sigma'}(x) = C(\nu) \frac{1}{\sigma'} \exp\left(-\left[\frac{|x|}{\sigma'}\right]^\nu\right) \quad (4.8)$$

Each coefficient is represented as a draw from a GGD with σ and ν calculated a priori or from the data itself.

The other model is a GMM, which represents the density of a subband as a sum of Gaussian distributions:

$$p(x) = \sum_{m=1}^M p_m f_m(x) \quad (4.9)$$

where $f_m(x)$ are $\mathcal{N}(0, \sigma_m^2)$, and $\sum_{m=1}^M p_m = 1$

A GMM with $M = 2$ typically results in a sufficiently close model to the observed density functions – typically one Gaussian will have a small σ to account for the peak at zero and the other a large σ to account for the tails. Thus, the probability is approximately the sum of (at least) two Gaussian densities with variances σ_m and mixture probabilities p_m .

The ML estimation using the above distributions has also been regularized using penalty functions based on the complexity of the resulting image[42]. The intuition is that real-world images tend to be fairly smooth. Complexity can be calculated in various manners, such as the length of the smallest binary string to code the image.

Moulin[43] has also shown the equivalence of hard thresholding and MAP estimation for certain statistical assumptions on the data.

While these models have yielded improved performance, they still are based on a statistical iid assumption of the transform coefficients in a subband, which ignores the persistence/clustering property. To accommodate a non-iid assumption, Crouse[44] implemented a Hidden Markov Model (HMM) based approach. The wavelet coefficients are modeled by a 2-state GMM, but the state variables (σ_j and p_j) are not assumed independent. By allowing these to be interdependent, either within the same subband (Markov chain) or across scales (Markov tree), the persistence/clustering properties can be modeled. As an example, from the magnitude decay property of the DWT, we can expect that the signal variances, σ_j^2 , should decay approximately exponentially across scales. Depending on the application, either or both of the inter/intrascale dependencies can be used.

The last approach I'll discuss is based on trying to exploit inter- and intrascale dependencies of the data[45] in a different manner. The algorithm is based on using a threshold, T_{sig} , to partition the coefficients into two sets, G_{sig} and G_{insig} based on magnitude. The process is started at the lowest (coarsest) scale in the decomposition. After the coefficients are segregated, they are processed separately, taking advantage of intrascale dependencies. The G_{sig} coefficients are experimentally seen to follow a Laplacian distribution with zero mean. Thus, the MAP estimate is a soft-threshold with threshold adapted to local signal variance[43]. The G_{insig} coefficients typically represent the smoother areas and are modeled as a GGD distribution. For this case the MAP estimator is a Wiener-type relation:

$$\hat{d}_{j,k} = \frac{\hat{\sigma}_{j,k}^2}{\hat{\sigma}_{j,k}^2 + \hat{\sigma}_n^2} d_{j,k} \quad (4.10)$$

where $\hat{\sigma}_{j,k}$ is an estimate of the wavelet signal variance and $\hat{\sigma}_n$ is the noise variance estimate. At each subsequent higher scale, G_{sig} and G_{insig} are based on the (denoised) magnitude of their parents, taking advantage of interscale dependencies (persistence).

The threshold T_{sig} is determined experimentally by maximizing the overall likelihood of the data based on the expected distributions of G_{sig} and G_{insig} . While this is a relatively simple and straightforward approach in the wavelet domain, the results are among the best reported.

One interesting note is that in both of the previous algorithms (and generally in compression algorithms), better performance was noted when exploiting interscale rather than intrascale (parent-child) relationships. Recently Liu has shown that this can be understood through looking at the Mutual Information (MI)[46]. She develops a methodology for estimating the mutual information between neighboring coefficients in the same band and across bands. The results show that the MI of the neighborhood is much higher than the MI of the parent. Also of note is that as the filter length increases, the parent-child MI tends to decrease. One possible explanation is that longer filters determine the coefficient from a larger neighborhood, 'diluting' the parent-child relationship. While the paper used only a single parent and multiple neighbors, the results do suggest a reason for the previous supremacy of intrascale models, with only slight benefits for adding interscale dependencies. This does not imply there is no interscale MI (which would be bad news for my super-resolution attempts), just that it is less than intrascale MI.

While the above work has all been based on signal independent AWGN, there has been some work on Poisson noise. The simplest approach is to work with the square root image under AWGN assumptions. In [47], Nowak provides an alternative, wavelet based approach, that uses multiple short exposure images and develops an estimator based on the cross-validation of the images, with slightly improved results over square-root processing. Nowak has also extended the HMM approach to account for Poisson statistics[48]. In [49], a new approach is used, based on multiscale multiplicative innovations. The concept is to use the framework of the Haar wavelet, where each coarser scale is the sum of the finer scale, but use multiplicative/ratio relationships instead. This allows a more precise model of the parent-child relationship

for Poisson data and the use of Bayesian prior. It also appears that an approach by Mihcak[50] should be extensible to Poisson data.

4.3 Image Restoration

Let us again take as a model the typical linear shift-invariant system:

$$g = h \otimes f + n \quad (4.11)$$

where g is the received image, f is the object, h is the PSF which we assume is known, n is the noise, and \otimes represents convolution. In the Fourier domain we have

$$G = HF + N \quad (4.12)$$

The optimal (in terms of MSE) linear solution for deconvolution with Gaussian noise is the Wiener filter, which is, in essence, an inversion of H followed by a regularization of the Fourier magnitudes based on the SNR. Donoho[36] proposed a new method using wavelets. The process is to deconvolve with H^{-1} , but then denoise in the wavelet domain. This has been termed the *Wavelet-Vaguellete Decomposition* (WVD). The benefit here is the superior performance of the non-linear wavelet domain denoising. The drawback is that AWGN, after inverse filtering, is no longer necessarily white. Thus, the denoising must account for this colored noise. Additionally, where H is zero, the noise variance after inverse filtering will tend to infinity.

To address this problem, Kalifa et al.[51, 52] proposed a *mirror wavelet basis*. The concept is to adapt the wavelet basis to a known zero at the digital frequency $\omega = \pi$. In particular, since diffraction induces a monotonic decrease in the frequency response, the only zero will be at the diffraction cut-off. To counteract the singularity in H^{-1} at the diffraction limit, a mirror wavelet basis is used. As an example, they use a 1-D signal with a singularity in the inverse filter at $\omega = \pi$. Figure 4.2 (solid line) shows the inverse filter frequency response for a circular aperture imaging system. The noise

variance, after inverse filtering, will be proportional to this value. With the standard DWT, the highest resolution subband contains data from the frequencies $\frac{\pi}{2} < \omega < \pi$. The extremely large noise variance near the singularity at $\omega = \pi$ will overwhelm any signal in this whole wavelet subband. This implies a need is to isolate the 'noisy' frequencies (due to inverse filtering) in the wavelet domain. The mirror wavelet basis is a wavelet packet transform where, in addition to iterating on the low-pass output, the high-pass output is also filtered as shown in figure 4.3. This yields the standard wavelet coefficients, $d_{j,k}$ as well as the mirror coefficients, $\check{d}_{j,k}$. The decomposition is called a mirror basis due to the symmetry of standard and mirror coefficients.

The notation d_j and \check{d}_j denote that these coefficients are at the same level of decomposition, but on opposite branches. The exception is the scaling coefficient signal a_{j-n} ($n = 2$ in figure 4.2), which has a wavelet coefficient counterpart which I'll denote as \check{d}_{j-n-1} .

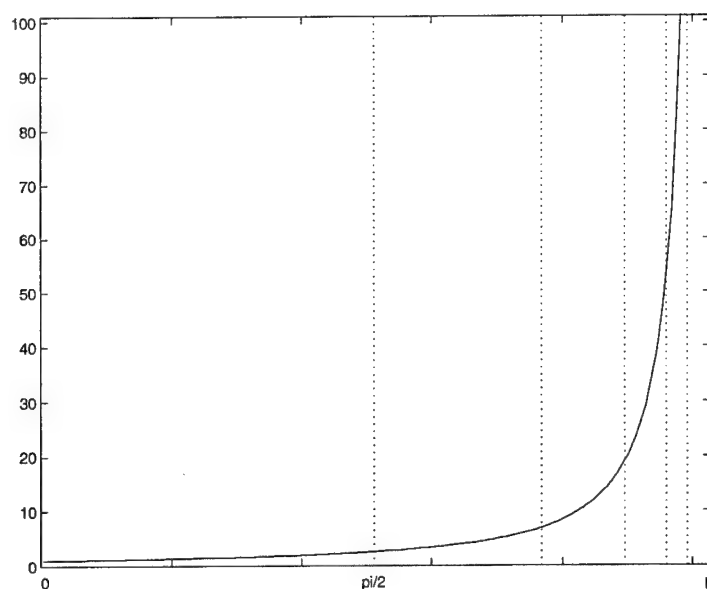


FIGURE 4.2. Frequency bands for mirror wavelet basis (dashed lines) and inverse filter for diffraction limited imaging system (solid line)

If we look in the Fourier domain, the frequency response of the wavelet basis, ψ_j ,

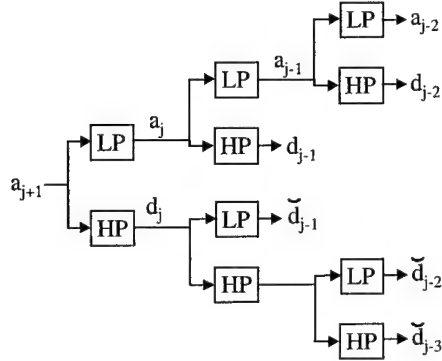


FIGURE 4.3. Algorithm for mirror wavelet basis decomposition

has energy concentrated on a interval

$$\left[\frac{\pi}{2^{J-j+1}}, \frac{\pi}{2^{J-j}} \right] \quad (4.13)$$

The mirror basis, $\check{\psi}_j$, will have energy concentrated around

$$\left[\pi - \frac{\pi}{2^{J-j}}, \pi - \frac{\pi}{2^{J-j+1}} \right] \quad (4.14)$$

as shown by dotted lines in 4.2. Taken together, the ψ_j and $\check{\psi}_j$ form an orthonormal basis. The significant benefit of this basis is that, for all subbands (except the highest frequency band), the noise variance remains within a bounded factor. This is the isolation of the singularity that is needed. Each of the individual subbands can be denoised with a subband dependent threshold based on the noise variance in that particular subband.

While powerful, the mirror wavelet basis requires a certain class of blurring operators, those with a single zero at $\omega = \pi$. In order to address the more general deconvolution problem, Neelamani[53] developed an approach by combining Fourier and wavelet domain regularization. The motivation is to use the benefit of non-linear wavelet denoising, but to account for the (possibly multiple and unknown a priori) zeros in H^{-1} . The concept is to perform a Wiener-like restoration first to regularize the the wavelet domain solution. The regularization in the Fourier domain is given

by

$$\hat{F} = \left(\frac{H^*(\xi, \eta)}{|H(\xi, \eta)|^2 + \alpha \frac{S_n(\xi, \eta)}{S_f(\xi, \eta)}} \right) G \quad (4.15)$$

where α is the regularization term. Using $\alpha = 1$ yields the standard wiener filter. As $\alpha \rightarrow 0$, the Wiener filter will approach a straight inverse filter, yielding a less blurred, but noisier image. When $\alpha = 0$ this is the straight inverse. For $0 < \alpha < 1$, the singularities in the inverse will be partially regularized. The intuition is to use Fourier regularization to reduce the noise amplification near zeros in the inverse filter, but still rely on wavelet denoising for the final solution. Optimizing the choice of α in a MSE sense is addressed, but must be set experimentally.

Additionally, there have been other variations based on various statistical assumptions and techniques. [42] uses a complexity measure to regularize the ML solution. [54] takes a stochastic approach and derives a Kalman filter for the coefficients. [55] develops a algorithm by iterating between calculations of the state coefficients in a 2-state GMM model and the MAP estimate of the coefficients. Complex wavelet packets have also been used[56]. This is based on using complex wavelets with a packet decomposition to denoise the result of applying the inverse filter, H^{-1} . The benefit appears to derive from the use of packets and the redundancy of the complex transform.

4.4 Image Interpolation

Wavelets can also be used for interpolation of images. Similar to the zero-padding common in the Fourier domain, we can zero-pad the wavelet-domain. This is accomplished by using the original image as the LL band and adding empty LH, HL, and HH subbands. In essence we create a new, higher scale without any data in it. We note that the algorithm for an inverse DWT upsamples the LL band signal, low-pass filters it, and then adds in the contribution from the higher scale wavelet coefficients.

If the higher scale wavelet coefficients are zero, there will be no contribution. Thus, this performs the same operation as classical interpolation - upsampling and filtering.

Wavelets yield a wide variety of possible 'filters' to use. Two limiting cases are 1) the Haar wavelet, which yields Zero Order Hold (ZOH), or replication interpolation, and 2) an infinite Sinc wavelet yields the traditional bandlimited interpolation.

In wavelet interpolation, if we have an estimation of the higher scale coefficients other than zero, the inverse DWT will add this information to the image. Remember this is due to the fact that wavelets coefficients contain only the difference information for the higher scale. This estimation of higher scale coefficients is the basis of the wavelet interpolation techniques below.

In [57], the basic premise is to estimate the higher scale data and then enhance it through a POCS-type algorithm. Assuming we are estimating the $j + 1^{st}$ scale, they start with the $j - 1^{st}$ scale since the j^{th} scale is thought to contain many non-meaningful extrema and be too noisy. The original estimate is formed by first identifying wavelet extrema at the $j - 1^{st}$ scale. These are then placed in the $j + 1^{st}$ scale using the extrapolated (exponential) relationship of the magnitudes of the extrema between scales. Once the extrema locations and magnitudes in the $j + 1^{st}$ scale are estimated, intermediate points are estimated through enforcing monotonicity[58].

The last step is to use three POCS constraints to improve the estimate:

1. The data must be consistent with wavelet transformed data, e.g. an element of V_{j+1} , the subspace of $L^2(\mathbb{R})$ which is the range of the wavelet transform.
2. A downsampled version of the interpolated data must equal the original data.
3. Local extrema in the higher-scale data should reflect sharp variations in the original data, i.e. extrema will persist among scales.

The first two are fairly trivial in projection operator: (1) is simply a forward and inverse wavelet transform of the data, and (2) is simple replacement of even values

in the upsampled (lower-scale) data with original data. The third constraint appears to perform the bulk of the interpolation, but is heuristic in nature. It is not very well described other than to enforce a penalty function for extrema different than the estimates or at locations far from the estimated location.

Carey et al.[59, 60] use the same concept of estimating the higher scale coefficients with a different approach. As above, they use the undecimated DWT and base their extrapolation on the $j - 1^{st}$ scale. However, they use local Holder regularity, a mathematical measure of image smoothness or regularity. Conceptually, they use the fact that low regularity areas (e.g. strong edges) have high correlation between scales and that the coefficients will decay exponentially between scales. Smooth or textured areas do not have such a relationship. By using a search to determine where coefficients have a exponential decay, they mark out the edges. The feature in scale $j - 1$ is reduced spatially to 1/4 its size (to reflect decrease in size for 2 scales higher) and copied into the $j + 1$ scale using a cubic spline approximation. Similar to above, the exact location is based on extrapolation from previous scale locations and the magnitude of the extrapolated data is adjusted, based on a mathematical fit of the data from previous scales. Results for Lenna show visual improvement of the image when magnified, compared to bandlimited interpolation. However, the Lenna results are significantly better than for Mandrill, apparently due to the relative proportion of sharp edges to texture. The algorithm searches only for edges that propagate through scales, and thus is best at images with such edges, and not for images with texture or other features.

The last technique is based on algorithms used in resolution enhancement of imagery[61], notably from NTSC to HDTV. The process is quite heuristic, but does show some minor improvement. The concept is to look for zero-crossings in the DWT coefficients which are assumed to be from edges. Higher resolution subband data is added where an edge is estimated based on statistics from a step edge, with the constants set by training images.

There are also two other interesting wavelet based techniques that demonstrate the use of wavelets in image processing and are similar in spirit to interpolation. The first concept is to perform image fusion in the wavelet domain. Several remote sensing systems (e.g. SPOT, Landsat) have both a high resolution panchromatic sensor and a lower resolution multispectral sensors. Several techniques have been developed to fuse the two products together and provide high resolution panchromatic imagery[62]. The overall technique is to either add or substitute some data from the higher scale wavelet transform bands of the panchromatic image into the wavelet transform bands of the multi-spectral data (creating them if necessary). The best success is from applying this addition/substitution to the intensity band of the data – i.e. transform into IHS form (intensity, hue, saturation), perform wavelet coefficient addition/substitution on the intensity band (leaving the H and S bands zero), inverse wavelet transform, and then transform back to the original multi-spectral bands such as RGB. The number of bands added/substituted is dependent on the relative difference in resolution of the panchromatic vs. multi-spectral images. The result is multispectral imagery with new higher scale data derived from the high resolution panchromatic image. This technique can be looked at as equivalent to wavelet based super-resolution of the multispectral data, given the a priori knowledge of the panchromatic data.

A similar technique has also been used in fusing multispectral imagery of a given resolution[63]. Given multiple bands of the same scene, often a goal is to develop a fused image. Averaging of all the data is one solution, but can wash out details that don't persist across bands. A simple wavelet-based approach that has shown success is to compute the various wavelet transforms and then average the coarse bands as these contain the coarse background information. For higher scale bands, choose the largest coefficient from any of the DWTs of the multispectral bands. Thus, the coarse data is consistent across the bands, and the details are picked from the band in which they are strongest.

CHAPTER 5

RADIALLY SYMMETRIC TRANSFORMS

5.1 Overview

Remembering that a circular aperture imaging system has a radial symmetry in the frequency domain, this chapter introduces a new transform that will maintain a similar symmetry. This will be very useful in deconvolution of imagery from circular aperture imaging systems. The first section lays out the intuitive motivation for the new transform, which is described in the second section. Following this, the transform is described and extended in a similar manner to the mirror basis described previously. The next sections discuss the actual deconvolution algorithm and the results. The final section is a short discussion on this transform and the curvelet transform.

5.2 Motivation

The mirror wavelet basis discussed above provides a powerful tool in image deconvolution. The power of this transform is the use of wavelets to denoise the results of an inverse filter. The mirror basis keeps the noise variance bounded in each subband (except, of course, the single subband at the singularity). However, this construction is based on a 1-D view of the frequency response. When the DWT is applied separably, the tiling of the frequency domain for a 2-D transform is shown in figure 5.1(a). Note that only a single quadrant (positive frequencies) is shown due to symmetry in the other three. As expected, the frequency $\omega = \pi$ is isolated in both the horizontal and vertical directions. The examples given by Kalifa et al. in [51, 52] use a separable blurring function, which isolates the zero in the frequency response of the filter at the edges of the frequency plot shown in 5.1(a), that is when either $\omega_x \approx \pi$ or $\omega_y \approx \pi$. However, if we look at a circular aperture imaging system, the

singularity in the inverse filter is located at a constant radial frequency, as shown in figure 5.1(b).

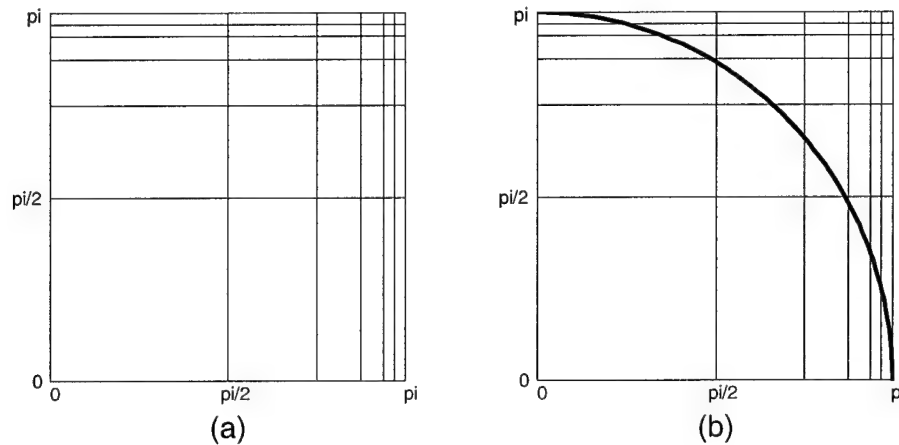


FIGURE 5.1. (a) 2-D frequency domain tiling of separable mirror wavelet basis, positive frequencies quadrant only, (b) with diffraction cut-off of circular aperture imaging system

Plotting this inverse filter frequency response along either frequency axis (ω_x or ω_y) is shown in figure 5.2(a), as discussed previously. However, in figure 5.2(b), the inverse filter is plotted along the diagonal. We see that the singularity is not isolated in the highest mirror wavelet subbands anymore. In fact, the singularity passes through the first mirror subband, i.e. the subband for frequencies $\frac{\pi}{2} < \omega_x, \omega_y < \frac{3\pi}{4}$. Thus, the separable mirror basis does not truly isolate the singularities of a circular aperture imaging system.

5.3 Radially Symmetric Transforms

In this section a new method to calculate a transform that will maintain radial symmetry in the Fourier domain will be developed. I will refer to this as a radially symmetric discrete wavelet-like transform (RS-DWT) where the symmetry refers to the Fourier domain. One item to note is that I have used the term ‘wavelet-like.’

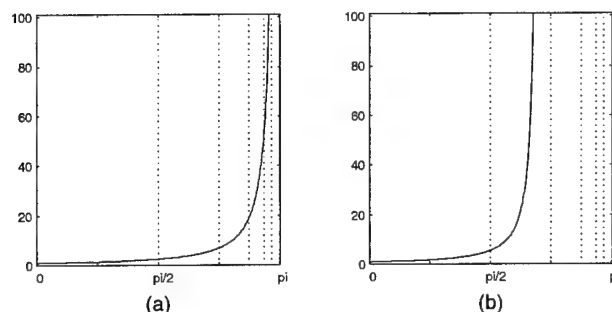


FIGURE 5.2. Frequency response of inverse filter for a circular aperture imaging system (solid line) along (a) either frequency axis and (b) the diagonal. Dotted lines denote the frequency bands of the mirror wavelet coefficients.

This is because the construction below will not, technically, be a wavelet transform. Others have developed a precise theory of multi-dimensional, non-separable wavelet transforms[64, 65], which lay out the criteria for multi-dimensional wavelets and discuss how to calculate them. However, the techniques are quite involved and often iterative in nature. My approach will be to take a simpler practical viewpoint, attempting to maintain the benefits stemming from a the wavelet transform while avoiding the complexities of a formal multi-dimensional wavelet transform. However, since the similarities to true wavelet transforms are quite strong, I will continue to use terms associated with wavelet transforms, although these must be understood to be applicable in a loose sense.

Since the 1-D wavelet transform can be described via convolutions with filters, h_0 and h_1 , it can also be described as a multiplication in the Fourier domain. Thus, the wavelet coefficients can be calculated via a Fourier-domain multiplication, where the appropriate multiplying mask in the Fourier domain is determined by a Fourier transform of the filter coefficients, with zero padding as necessary. In general, this will require a complex representation of the filter frequency response H_0 . To allow use of amplitude only values, the filter must be zero phase (or generalized linear phase by taking into account the implied shift). Such a constraint on the Fourier transform

requires symmetric filter coefficients. The only finite-length symmetric orthogonal wavelet filter is the Haar basis. In order to allow flexibility beyond the Haar basis while maintaining amplitude only masks will require the use of biorthogonal wavelets. Many zero-phase biorthogonal wavelets have been developed, and these will be used.

In order to simplify the discussions initially to enable the concept to be readily understood, assume use of the Sinc wavelet system[32, p. 60]. The Sinc wavelet filter is given by

$$h_0(n) = \text{sinc}\left(\frac{\pi}{2}n\right) \quad (5.1)$$

for a signal sampled at the Nyquist frequency. This filter is obviously an IIR filter and thus cannot be implemented in the spatial domain. This example is instructive, however, as it provides a simple and intuitive introduction of the new concepts. The Sinc function in the spatial domain corresponds to a Rect function in the Fourier domain. Thus the implementation of this filter is multiplication with a Rect function in the Fourier domain. To implement a 1-D DWT we can then use these binary masks (since Rect is 0 or 1) in the Fourier domain. The extension to a 2-D separable transform can be accomplished as individual 1-D transforms on the rows/columns.

Alternatively, we can reformulate the individual row/column operations into a single multiplication in the Fourier domain. If we look at an individual pixel in the Fourier domain, it will be multiplied by the applicable frequency response from the row and column filters. We can combine this into a single 2-D multiplication using the outer product of the filter responses,

$$W(\xi, \eta) = h_r(\xi) h_c(\eta) \quad (5.2a)$$

$$\text{or } W = h_r h_c^\dagger \quad (5.2b)$$

where W is the frequency mask, h_r and h_c are the row/column filters.

It is also important to note that there will be four subbands (LL, LH, HL, HH), and thus four masks need to be used – each of which can be calculated from the outer product of the respective low-pass or high-pass Rect functions. The masks for the

four subbands are shown in figure 5.3, where one is represented by white, and a zero by black. The wavelet coefficients are given by, for the example of the HH band,

$$d_j^{HH} = \mathcal{F}^{-1} \{W_{HH} * \mathcal{F} \{d_{j+1}\}\} \quad (5.3)$$

where W_{HH} is the respective frequency mask. In this way, we can calculate the complete transform representation.

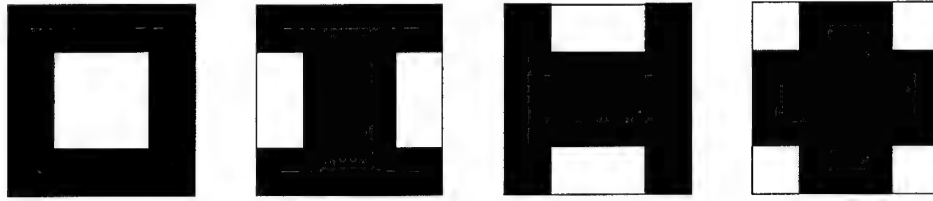


FIGURE 5.3. The masks used for the (left to right) LL, HL, LH, HH subbands

If I restrict attention to only the undecimated transform, aliasing will not be an issue. Thus, it is possible to combine the three high-pass subbands (LH, HL, and HH) into a single band. This yields a single high-pass (wavelet) image and a single low-pass (scaling) image at each scale. The masks for 3 levels of wavelet decomposition and the resulting scaling image are shown in the top row of figure 5.4. An example of this transform applied to Lenna is shown in the bottom row. Note the ringing in the wavelet images. This is expected due to the nature of the Sinc wavelet – increased localization in the frequency domain implies lack of localization in the spatial domain.

This method of calculating the DWT yields the undecimated DWT coefficients. The decimated form may be determined by downsampling the result. Note however, that the first level of decomposition must be downsampled by a factor of two (in each dimension) whereas the second must be downsampled by a factor of 4, etc.

This concept of frequency masks can then be extended and used to enforce radial symmetry in frequency space. For each point in the frequency domain, we can calculate the radial frequency ($\rho = \sqrt{\omega_x^2 + \omega_y^2}$), and assign to that point the frequency

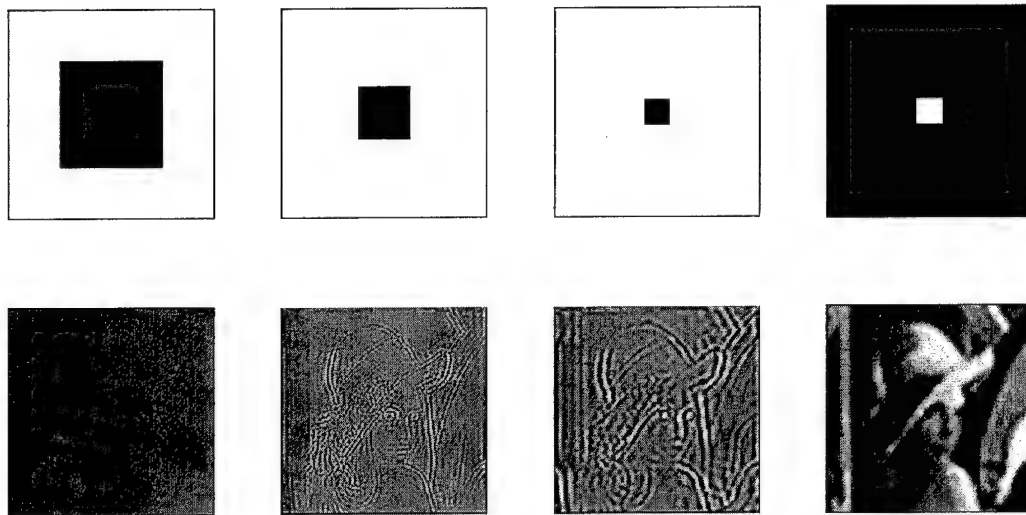


FIGURE 5.4. Top row is the first three wavelet masks and the resulting scaling function frequency mask for the *sinc* wavelet. Bottom row shows the corresponding wavelet/scaling images.

response of a 1-D wavelet at the given frequency. This will enforce a radially symmetric filter. One issue is that a 1-D filter will have a filter response out to $\omega = \pi$. With a rectangular sampling grid, radial frequencies out to $\omega = \sqrt{2}\pi$ will be present in the ‘corners’ of the Fourier domain representation. This issue can be addressed in several ways:

- Assigning all pixels with $\omega > \pi$ to be the value at $\omega = \pi$.
- Using a first decomposition scale to represent the data in the ‘corners’ by removing all frequencies $\omega > \pi$.
- Assume periodic replication and fold the frequency response of the filter around $\omega = \pi$.

Since we are assuming a critically sampled circular aperture system, there is no information content above $\omega = \pi$. Thus, the third option is not preferable as it would introduce noise artifacts from frequencies above the diffraction cut-off into the high-

and low-pass bands. The first and second options make the most sense. The second option will be used when not otherwise noted, although little significant difference was noted in performance between the first two. The first option simply adds a small amount of additional noise into highest wavelet subband, since critical sampling is assumed. The second option is equivalent to multiplication in the Fourier domain with a large *circ* function, which is equivalently a convolution with a small Bessel-like PSF.

If we use the same 1-D Sinc-wavelet transform, the 2-D RS-DWT corresponds to a Circ function in the Fourier domain or loosely speaking, a 2-D *jinc*-function wavelet. Following the above approach, we use binary masks in the Fourier domain as shown in the top row of figure 5.5. Each successive scale is a dyadic frequency band. An example of the resulting transform is shown in the bottom row of figure 5.5. Note that the first scale contains only the ‘corner’ data discussed above. A loose interpretation of this transform is that it approximates an undecimated DWT using *jinc*-function wavelets, whose radial symmetry is similar to that of circular aperture imaging systems.

Note that due to the combination of the high-pass bands and the radial symmetry, there is not a way to express this transform in a decimated form. The low-pass subband in the binary mask case could be decimated by a factor of two in each dimension, without concern about aliasing. However, any FIR wavelet will not be perfectly band-limited below $\omega = \frac{\pi}{2}$, and thus aliasing will occur in downsampling. This aliasing will not be canceled by reconstruction since only the low-pass band was downsampled. The high-pass band cannot be downsampled without loss of information (it contains the rough equivalent of the HL, LH, and HH subbands).

While the above cases demonstrate the use of Fourier masks, the ringing in the transform data (due to the Sinc’s frequency response) is not desirable. In fact, many of the benefits of wavelet transforms stem from the spatial locality of the transform. However, it is a simple extension of the above techniques to use any FIR wavelet. The

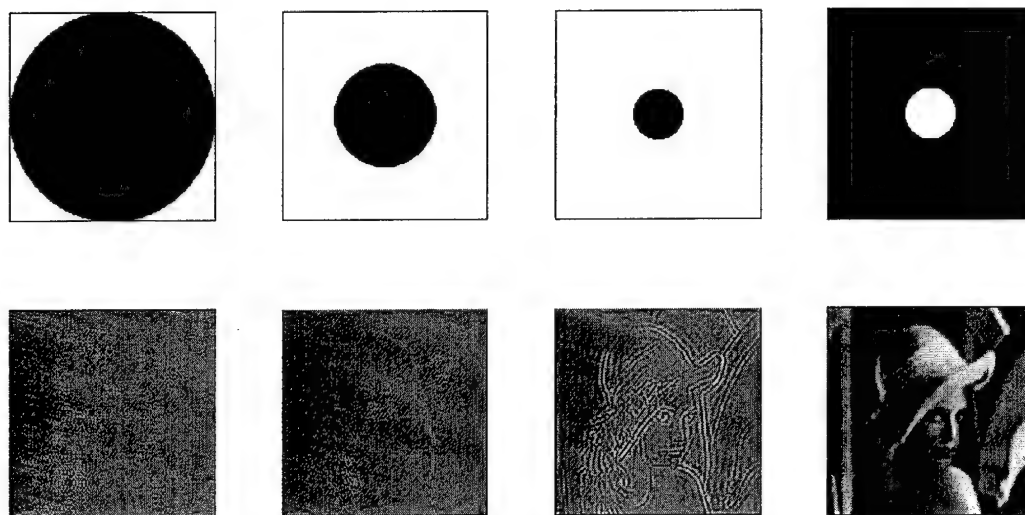


FIGURE 5.5. Top row is the first three wavelet masks and the resulting scaling function frequency mask. Bottom row shows the corresponding wavelet/scaling images.

concept is to use the frequency response of a 1-D wavelet filter to generate ‘gray-level’ masks. One additional complicating factor not necessary to include above (due to the binary nature) will be that since undecimated transforms are assumed, the filters must be upsampled. This will lead to periodic lobes in the frequency response masks due to the periodic replication of the filter response. This was previously shown in section 3.5 and the impact discussed there.

The main-lobe only masks when biorthogonal 9/7 wavelet filters are used are shown in the top row figure 5.6. These show where the majority of the energy in a subband will be located. The next row shows the actual masks used that include the periodic lobes from the filter upsampling. The last row shows the wavelet and scaling coefficient images. Note that the ringing artifacts are not present, as expected, while a radial frequency symmetry and perfect reconstruction is maintained. Any 1-D plot through the center of the Fourier domain masks will be identical to the 1-D frequency response of the wavelet used.

The question remains as to what extent does this transform capture the critical

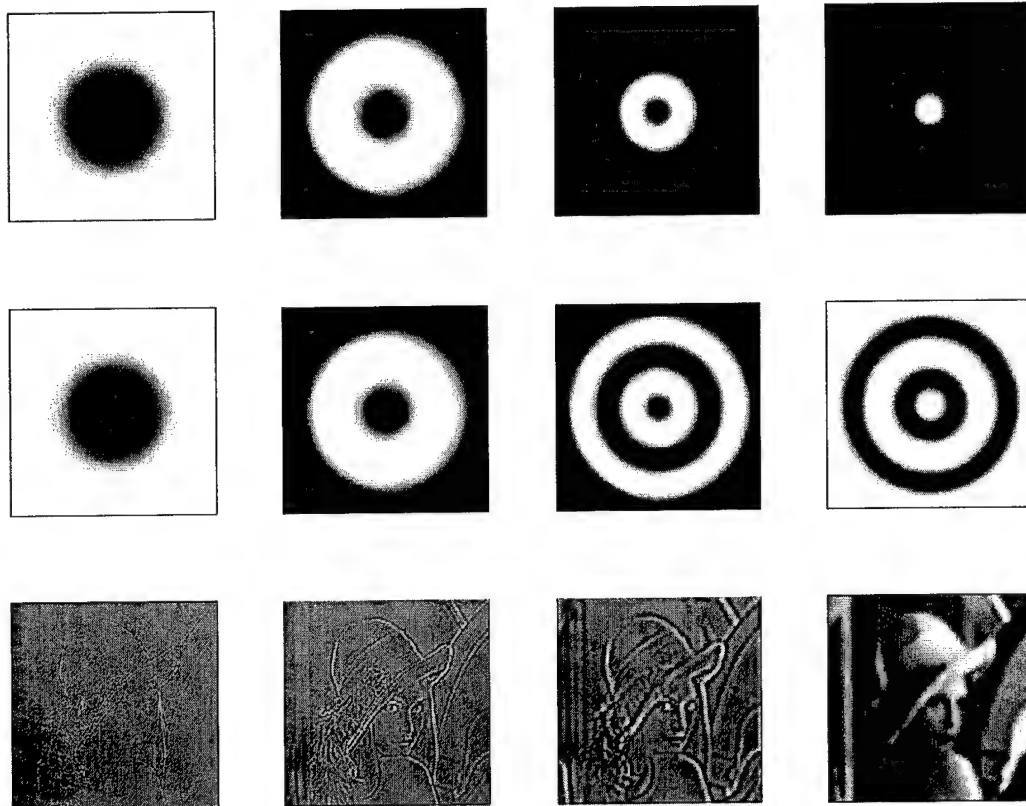


FIGURE 5.6. Top row is the first three wavelet masks and the resulting scaling function frequency mask for the bi-orthogonal 9/7 wavelet pair. Middle row shows the same masks with applicable periodic lobes. Bottom row shows the corresponding wavelet/scaling images.

properties of the DWT. In order to provide an answer, I will provide an intuitive argument based on the transform itself demonstrating how it meets the properties listed in section 3.4.

P1: Locality: Since the transform is a multiplication in the Fourier domain, we can calculate an equivalent convolution kernel in the spatial domain and examine its spatial extent. As shown in figure 5.7, there is a sharp decay in the magnitude of the kernel away from the origin. In the figure, a 1-D slice through the center of the 2-D high-pass kernel is shown in both linear and semi-log formats. Note

that the amplitude is shown on top, while the logarithm of the magnitude is shown below. There is a sharp drop-off of 4-5 orders of magnitude within 8 pixels of the center, whereas the 1-D wavelet used is the biorthogonal 9/7 pair, with length 7 for the initial high-pass. Over 99% of the energy in the 2-D RS-DWT filter is concentrated in the area within 8 pixels of the center.

P2: Multiresolution: The transform images provide insight into images properties across a set of scales. This is obvious from both the designed frequency response of the transform and the resultant set of transform images (e.g. figure 5.6, last row).

P3: Sparse Representation: In order to determine sparsity of the transform, I will use two metrics: 1) a plot of the percentage of the total transform energy represented in a set percentage of the coefficients, and 2) estimated probability density functions. The top plot of figure 5.8 shows the first metric. The solid line shows the result for the RS-DWT while the dotted line is the undecimated DWT. The results are almost identical. The bottom two plots show a histogram of the wavelet coefficients and the best fit to a GGD distribution. The middle one is the RS-DWT, while the lower is the undecimated DWT. The results are very similar, with low relative entropy between the two. For comparison, figure 5.9 shows the same type of plots as in figure 5.8, but now compares the RS-DWT with an ideal band-pass (rect function filter) image covering the same frequency band. Note the differences between the two, not only in the top plot, but especially the change in GGD exponent (shape parameter) and increase in relative entropy. Thus, the RS-DWT does a comparable job to an undecimated DWT at sparsely representing the data. Similar results were noted with a range on natural imagery.

P4: Decorrelation: The RS-DWT maintains the approximate decorrelation of coef-

coefficients as seen in the undecimated DWT. Figure 5.10 plots the correlation coefficient versus pixel separation for both the RS-DWT and undecimated DWT. The results are extremely similar, lending support to the near decorrelation of coefficients.

5.4 Modified mirror wavelet basis

Using the above transform as the beginning, we can extend to a mirror wavelet packet basis while maintaining radial symmetry. Due to the RS-DWT, each decomposition level yields a single high-pass and a single low-pass band which can be further decomposed as desired. A graphical outline of the transform used is shown in figure 5.11 where the biorthogonal wavelet filters used are h_0 and g_0 for the low-pass filter and h_1 and g_1 for the high-pass filters. Several notes on the implementation are necessary.

- At each scale, the filter is modified by the insertion of $(J - 1 - scale)$ zeros between each filter coefficient (upsampling).
- The low-pass and high-pass filters are swapped at every other scale in the transform for the mirror packet coefficients. This is a direct result of the upsampling. h_0 , when upsampled by a factor of 2 (one zero between each coefficient) will have the primary band-pass of interest at $\frac{3\pi}{4} < \omega_p < \pi$, which is actually a high-pass for the scale $J - 1$ high-pass output.
- At each level, the 2-D Fourier mask used is derived from the 1-D frequency response of the respective filter using the radial symmetry method discussed above.
- The inverse transform is calculated as expected, using h_1 and g_0 for the filters.
- This is a perfect reconstruction (PR) filtering operation.

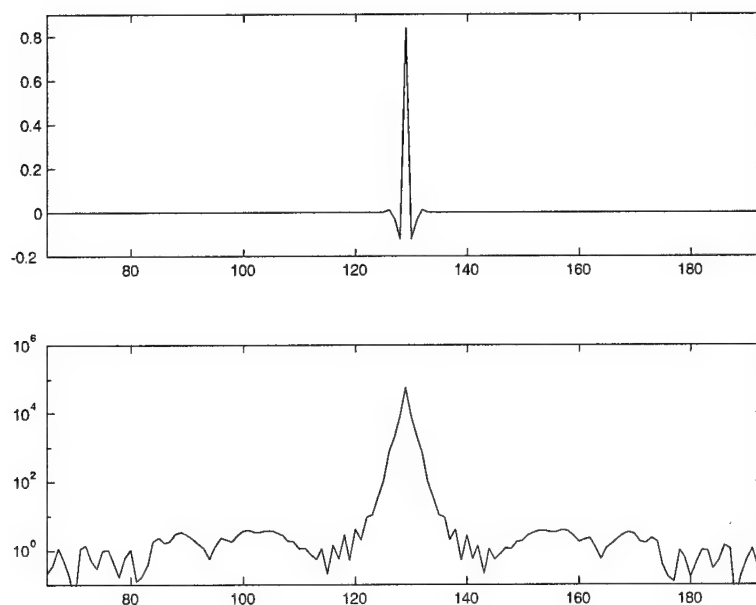


FIGURE 5.7. 1-D plot across 2-D kernel of HH band (top) and log plot to show decay (bottom). The x-axis is pixels, while the y-axis is relative amplitude.

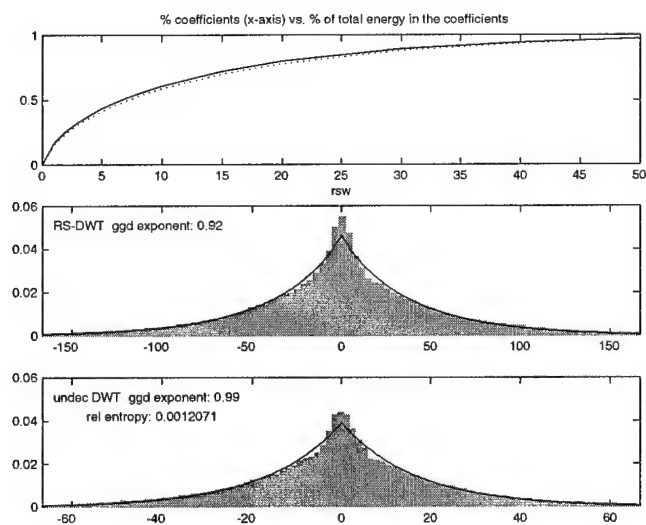


FIGURE 5.8. Plots for measurement of sparsity of transform representations comparing RS-DWT and undecimated DWT. In top plot RS-DWT is solid and undecimated DWT is dotted.

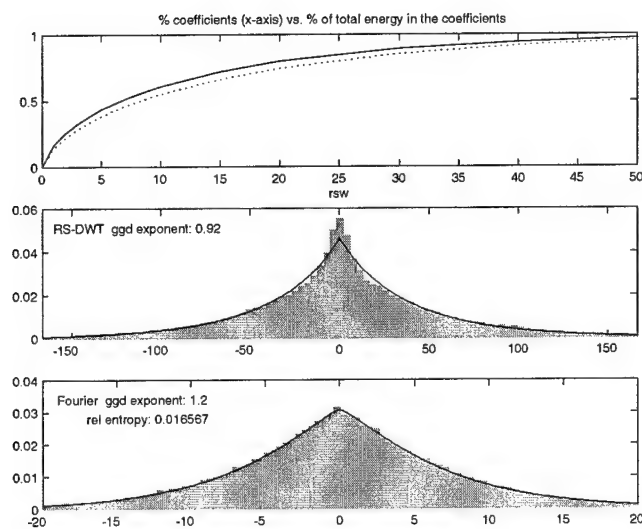


FIGURE 5.9. Plots for measurement of sparsity of transform representations comparing RS-DWT and ideal band-pass image of same frequency band. In top plot RS-DWT is solid.

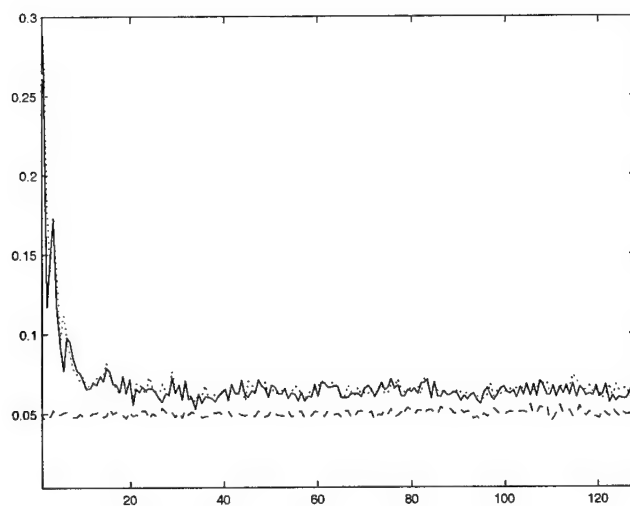


FIGURE 5.10. Correlation coefficient of wavelet coefficients for RS-DWT (solid line) and undecimated DWT (dotted line) plotted vs. pixel separation. For comparison, the correlation of a white Gaussian noise image is shown as a dashed line.

The first two notes discussed above are the result of implementing the algorithm à trous. The resulting frequency plot in 2-D (positive frequencies quadrant) is shown in figure 5.12. As desired, the singularity in the inverse filter of a critically sampled imaging system is now isolated by the mirror packet subbands.

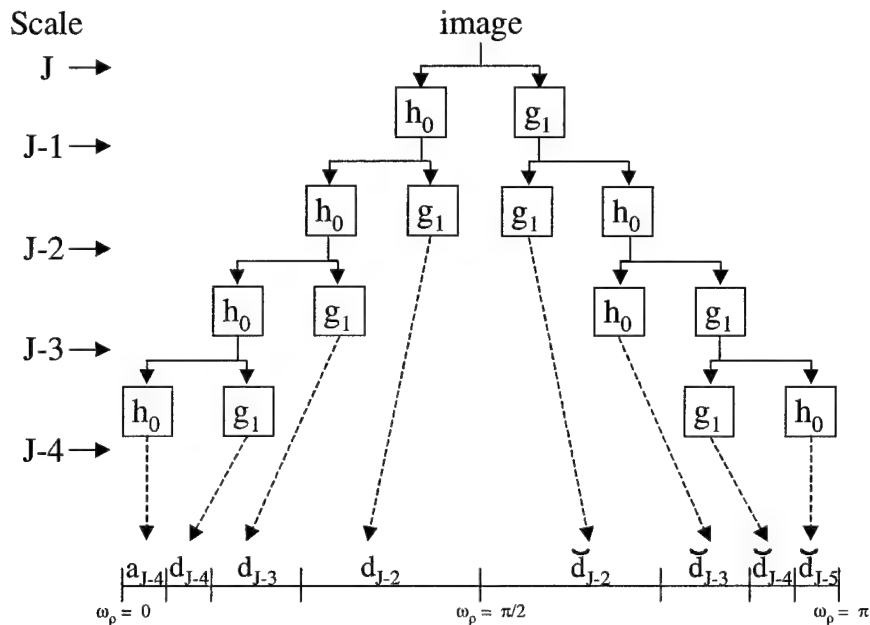


FIGURE 5.11. Graphical algorithm for calculating the RS-DWT mirror basis transform.

5.5 Discussion of Image Restoration Algorithm

5.5.1 Algorithm overview

This section will provide an overview of the proposed algorithm, with subsequent sections further discussing the details.

To compute the simulated input image, an image is blurred using a critically sampled circular aperture OTF (diffraction cut-off at $\rho = \sqrt{\omega_x^2 + \omega_y^2} = \pi$) and AWGN of a given variance is added. The degraded image is the only input (other than

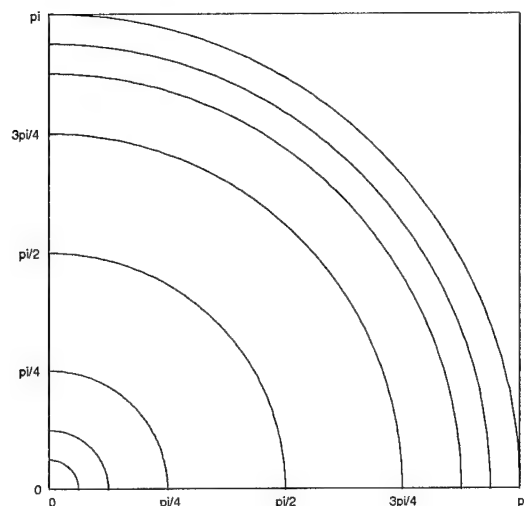


FIGURE 5.12. Frequency plot of the subbands in a RS-DWT mirror basis. Note that only positive frequency quadrant is shown.

knowledge of the blurring function) to the restoration algorithm. The first step is to estimate the noise variance and form a new image consisting only of white Gaussian noise of the estimated variance. After this, both the input data and the noise image are inverse filtered with the blurring operator and forward transformed using the RS-DWT. From the noise only image an estimate of the subband dependent noise variance is computed. Using these values, the transformed input image is then denoised, either via soft thresholding or a Wiener filtering of the wavelet coefficients. An inverse RS-DWT will then yield the result. Figure 5.13 outlines the process.

5.5.2 Noise variance estimation

The key element of the algorithm is denoising of the data in the RS-DWT domain. Thus, a thorough understanding of the noise properties in this domain is required, particularly the variance as it is used to set the threshold for denoising. First, an accurate estimation of the AWGN present in the original image is necessary. Since an orthogonal transform does not change the noise properties, the highest scale subband

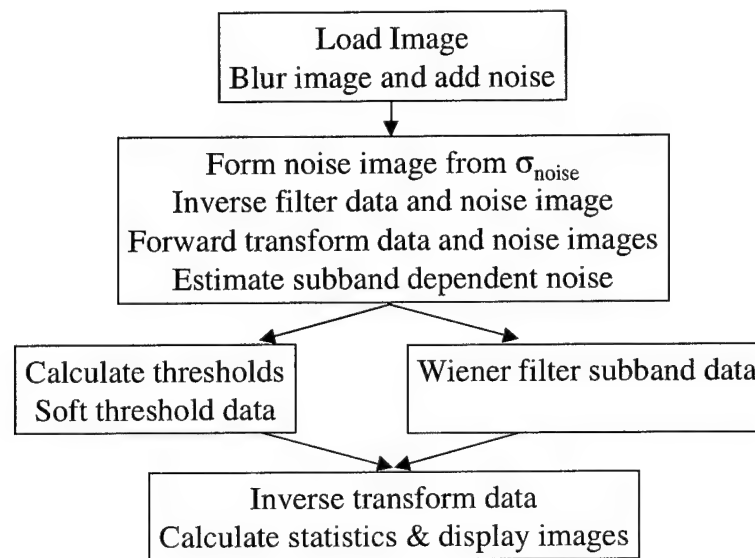


FIGURE 5.13. Algorithm overview for RS-DWT based deconvolution.

in an orthogonal DWT will have the same noise variance as the original data. Using the median estimator ($\sigma = \frac{\text{median}}{0.6745}$) on these DWT coefficients, the result is within a few percent for most natural images. The key is that natural images have less signal energy in the higher frequencies, allowing better estimation of the noise using the median. By using an undecimated, separable mirror-packet transform which isolates the highest frequencies, I have been able to reduce the error to below a percent. Using the variance estimated by this method has no noticeable change in deconvolution performance than using the actual value.

Next, an estimate of the noise in RS-DWT domain after inverse filtering is required. This is the image that must be denoised to arrive at the restored image. Even though the input image has AWGN, the inverse blurring operation will significantly color the noise. Additionally, the RS-DWT (a non-orthogonal transform) will also tend to color the noise further. While both of these ‘coloring’ effects are deterministic in nature, I estimate the noise variance in a given subband directly. However, this cannot reliably be accomplished in the presence of the signal. The concept im-

plemented is to form a noise only image with white Gaussian noise of the estimated variance. Inverse filtering and transforming this image to the RS-DWT domain will yield a noise spectrum which is representative of that in the transformed input image. Using the median estimator once again yields an estimate of the subband dependent noise variance, $\sigma_{j,noise}^2$, for scale j .

5.5.3 Denoising approach

Two denoising approaches were used. In the first case, $\sigma_{j,noise}$ can be used to calculate a threshold to denoise the data. While a constant multiplier to the standard deviation was used to start, it was noted that the error images (image formed as the difference between the true and restored image) had significant energy in the higher frequencies. If the threshold was increased to diminish this noise, low frequency errors were introduced. Thus an adaptive multiplier was used, one that increased with the inverse filter frequency response. The rationale can be understood by looking at the impact of the threshold in a deconvolution-type denoising. For simple AWGN denoising, the cost (in restored image degradation) of not thresholding a coefficient 'mostly' due to noise is constant in each subband – a result of Parseval's equality for the orthogonal case. For the biorthogonal case, it is still approximately true with the approximation better for tighter frames. However, when an inverse filter is involved, the cost changes. For high frequencies, the inverse filter frequency response will be large. Thus, if noise is passed through the soft-thresholding, it will be significantly amplified noise. Thus, at higher frequencies we want to be more robust and threshold out more of the noise, possibly at the expense of some signal energy, to prevent large noise in our reconstruction. However, the lower subbands can tolerate a lower threshold that might allow more of the noise through, but also more of the signal energy.

The second approach is to Wiener filter the wavelet coefficients. Note that this

is not a Wiener-Helstrom deconvolution, but simply a denoising of the wavelet coefficients in an MSE optimal sense, assuming Gaussian noise. In general this is given by:

$$\hat{d}_{j,k} = \left(\frac{\delta_{j,k}^2}{\delta_{j,k}^2 + \sigma_n^2} \right) d_{j,k} \quad (5.4a)$$

$$= \left(\frac{1}{1 + \frac{\sigma_n^2}{\delta_{j,k}^2}} \right) d_{j,k} \quad (5.4b)$$

where $\delta_{j,k}$ is the true wavelet coefficient of the object, $d_{j,k}$ is the wavelet coefficient of the transformed data, and $\hat{d}_{j,k}$ is the new estimate. In fact, it was shown[66] that hard thresholding can be interpreted as a Wiener filter for an extreme model of $\delta_{j,k}$ given by

$$\delta_{j,k} = \begin{cases} \infty, & |d_{j,k}| > T \\ 0, & else \end{cases} \quad (5.5)$$

While this is the MSE optimal denoising solution, it requires a model for the coefficients we are trying to estimate, the $\delta_{j,k}$. The solution is to restore the image using soft-thresholding as discussed above. This restored image can then be used as the estimate of the coefficients in a second (and different) wavelet basis, as shown in figure 5.14. This approach did improve results, but tended to over-smooth the result, leading me to regularize the Wiener filtering

$$\hat{d}_{j,k} = \left(\frac{1}{1 + \alpha \frac{\sigma_n^2}{\delta_{j,k}^2}} \right) d_{j,k} \quad (5.6)$$

where the α term reduces the amount of ‘smoothing’ performed by the filter.

5.5.4 Wavelet basis

The last algorithm detail to be discussed is the choice of RS-DWT mirror packet basis. First is choice of the 1-D wavelet family. For the results shown below, the 9/7 biorthogonal filter pair[67] and the Cohen-Daubechies-Feauveau (CDF)[68] family of

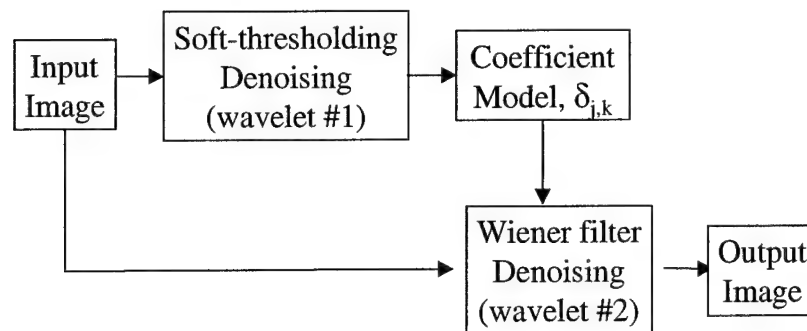


FIGURE 5.14. Wiener filter denoising algorithm

filter pairs were used. These were chosen as they have been shown to perform image compression well for a wide class of imagery. Next, the level of decomposition is required. It is theoretically possible to decompose an $n \times n$ image into $\log_2 n$ levels. However, this is not necessarily the optimal choice for a particular application, either in terms of performance or computationally.

For standard denoising applications, multiple levels of DWT decomposition are used to provide separation of the low-pass (scaling coefficients) from the wavelet coefficients. This is necessary as the scaling coefficient data is not denoised. Thus it is important to transform to a level at which the scaling coefficient data is fairly reliable estimate of the true data. Since the scaling data is the result of multiple low-pass filters, the noise tends to be smoothed out after multiple levels. I haven't found any formal study on performance vs. number of levels of decomposition, but typically at least 4-5 transform levels are used.

For the current work, the mirror packet transform used is shown in figure 5.12, where both the high and low frequencies were decomposed 4 times. While not necessarily optimized, further decomposition did not noticeably improve performance for the class of images tested, while computational speed dropped quickly. An additional aspect to consider is that due to the growth in inverse filter, at a certain frequency, all object information will be obscured by noise. This point is dependent on the noise

in the image, but also points out that further decomposition (at the high frequencies) will not necessarily improve performance.

5.6 Results

As a baseline, all results refer to images which are scaled in 0 - 255. For display below, I used the Lenna and Urban image shown in figure 5.15. For comparison purposes, the results will be compared with Wiener filtering. Two different Wiener filters will be used. The first will be what I'll call a fair Wiener filter, meaning the input to the Wiener filter will be only the degraded image and knowledge of the blurring function. When possible, I'll use the iterative signal PSD estimation discussed above, as implemented in [69]. I will also present the results for an 'unfair' Wiener filter which will be given perfect knowledge of the signal and noise PSD. This is given for comparison purposes as it is the MSE optimal solution for an LSI restoration under AWGN. This is unfair knowledge since the exact signal PSD is never known in practice. All of the results for the RS-DWT algorithm use only the degraded image and blurring operator as a priori knowledge.

The proposed algorithm was used to restore a degraded Lenna image with $\sigma_{noise} = 1$. The results on part of the image, enlarged to show detail, are shown in figure 5.16. The top row, (a) and (b), show the original and degraded image. (c) and (e) are the results of the fair and unfair Wiener filtering, respectively. (d) is the result of the RS-DWT using soft-thresholding; (f) is the result when wavelet domain Wiener filtering is added. Notice that all four of the restorations do a decent job of restoring the image details. However, the Wiener filter results show more noise remaining in the smooth areas of the imagery, notably the face. Even the unfair Wiener filter is less smooth than the RS-DWT results. This is typical of results here and in other wavelet domain restoration attempts - the wavelet domain retains the ability to capture sharp transitions and smooth areas, while the Wiener filter, in order to

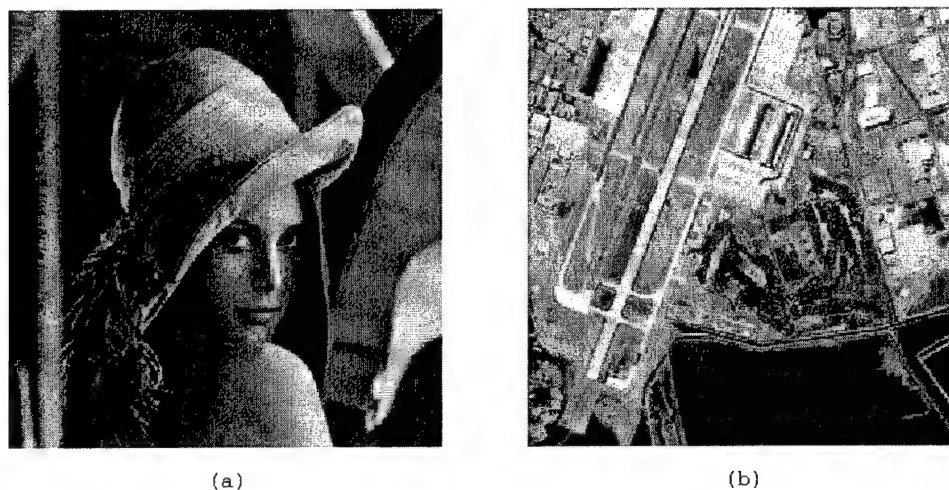


FIGURE 5.15. Images used in restoration. (a) is Lenna, (b) is Urban.

capture sharp edges usually leaves a noisy look in smooth areas. The results in terms of ISNR are shown in table 5.1. Note that the ISNR values are based on the whole image, not just the enlarged part displayed in the figures. In this case, the RS-DWT actually outperforms the unfair Wiener filter in terms of ISNR by about a third of a dB. For results observed across multiple images, it was noticed that while ISNR for the unfair Wiener filter and RS-DWT deconvolution were comparable. Note that the images were individually scaled for display.

Next the algorithms are used on the Urban image, with $\sigma_{noise} = 1$ again. The results shown in figure 5.17 are similar to that for the Lenna image above. The details, such as aircraft on the ramp were reconstructed fairly well by both algorithms, but the RS-DWT was noticeably smoother in smooth areas of the image. Figure 5.18 shows another section of the same image and restorations. Notice the road near the top and the wire across the water are restored, but again the RS-DWT result is much smoother. Figure 5.19 displays a 1-D horizontal plot near the bottom of the image across the water. The plots, which are offset from one another for visibility are,

from bottom to top, the original image, the degraded image, the fair WF estimate, the unfair WF estimate, and the RS-DWT estimate. Again, this demonstrates the ability to reconstruct sharp transitions, such as around pixels 170 or 450, while also keeping smooth areas, such as the water between roughly pixels 200 and 400. The ISNR is over 2 dB better than the fair WF case, and about a third of a dB lower than the unfair Wiener filter results.

The next set of results are for the Urban image with more noise, $\sigma_{noise} = 4$. Figures 5.20 and 5.21 show the results for similar sections of the images as before. Once again, the RS-DWT tends to be smoother in the smooth regions, although there are artifacts that are starting to appear in the RS-DWT results as well. These artifacts are due to noise coefficients that have not been thresholded, which results in an artifact of the wavelet kernel in the restored image.

Image	Noisy Image SNR	Wiener (fair) ISNR	Wiener (unfair) ISNR	RS-DWT (soft-thresh) ISNR	RS-DWT (wavelet WF) ISNR
Lenna $\sigma = 1$	30.62	5.37	6.45	6.26	6.78
urban $\sigma = 1$	25.66	2.85	5.82	5.04	5.47
urban $\sigma = 4$	24.91	-0.36	2.63	1.87	2.06

TABLE 5.1. RS-DWT deconvolution results

The results above are based on using the 9/7 biorthogonal wavelets for threshold denoising, and the CDF 13/5 wavelets for Wiener denoising. Results were similar when either the 9/7 or CDF wavelets were used. However, there was a performance difference noted depending on which filter of the biorthogonal pair was used as the low-pass synthesis filter. Around a third of a dB improvement was seen when the low-pass synthesis filter was the longer of the two.

5.7 Relationship to Curvelet Transforms

Recently, Candes and Donoho have introduced curvelets as a transform to represent 2-D images that are smooth except for edges which are smooth curves[70]. In [71], Do and Vetterli show that a pyramidal directional filter bank (PDFB) yields a curvelet-like transform. A PDFB is a transform made up a directional filter bank, as designed in [72], with the addition of a multiscale decomposition. Graphically, this is demonstrated in figure 5.22 which shows the directional and multiscale decompositions separately in (a) and (b), and the combined transform in (c). In practice, they are implemented independently – the high-pass output of the multiscale transform is further transformed by the directional filter bank. When the number of directions in the directional decomposition is doubled at every other scale, the transform satisfies several key properties of the curvelet transform.

Do and Vetterli rely on the Laplacian Pyramid to perform the multiscale decomposition, due to the fact that it provides a single high-pass and single low-pass output from a 2-D transform, as opposed to the 4 subbands from a separable DWT. Using this transform, much more detail is retained in denoising of imagery than by wavelets alone. The RS-DWT above is also another way to achieve the single high/low-pass output, and would be a natural candidate for the multi-scale decomposition. This is especially true for problems where the radial symmetry is important, such as in colored noise or deconvolution. It is also reasonable to expect the wavelet-like properties of the RS-DWT to improve performance over a Laplacian Pyramid.

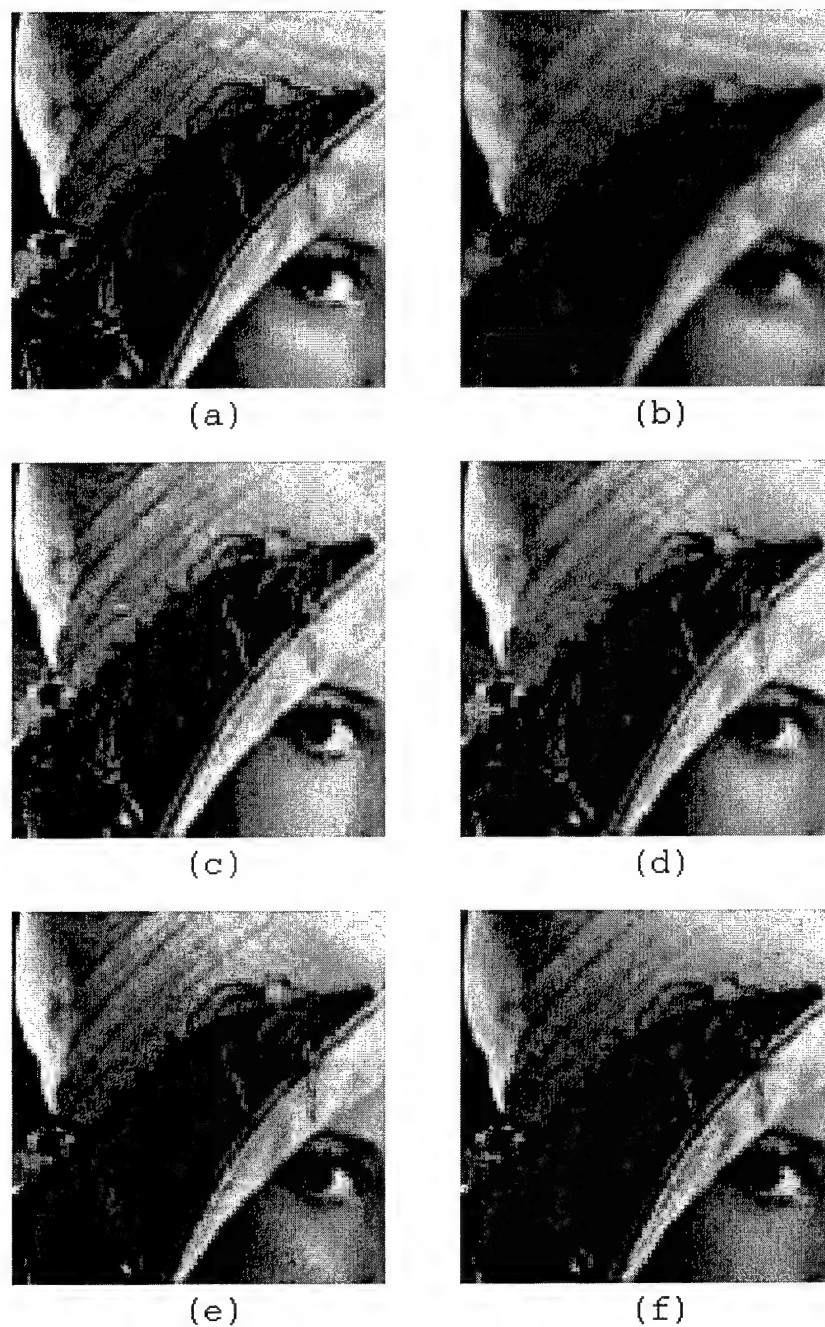


FIGURE 5.16. Restoration results. Enlarged section of image to show detail. (a) is original, (b) is degraded image (SNR 30.62), (c) results of fair WF (ISNR 5.37), (d) results using RS-DWT (ISNR 6.26), (e) results of the unfair WF (ISNR 6.45), (f) results using RS-DWT with wavelet domain WF (ISNR 6.78).

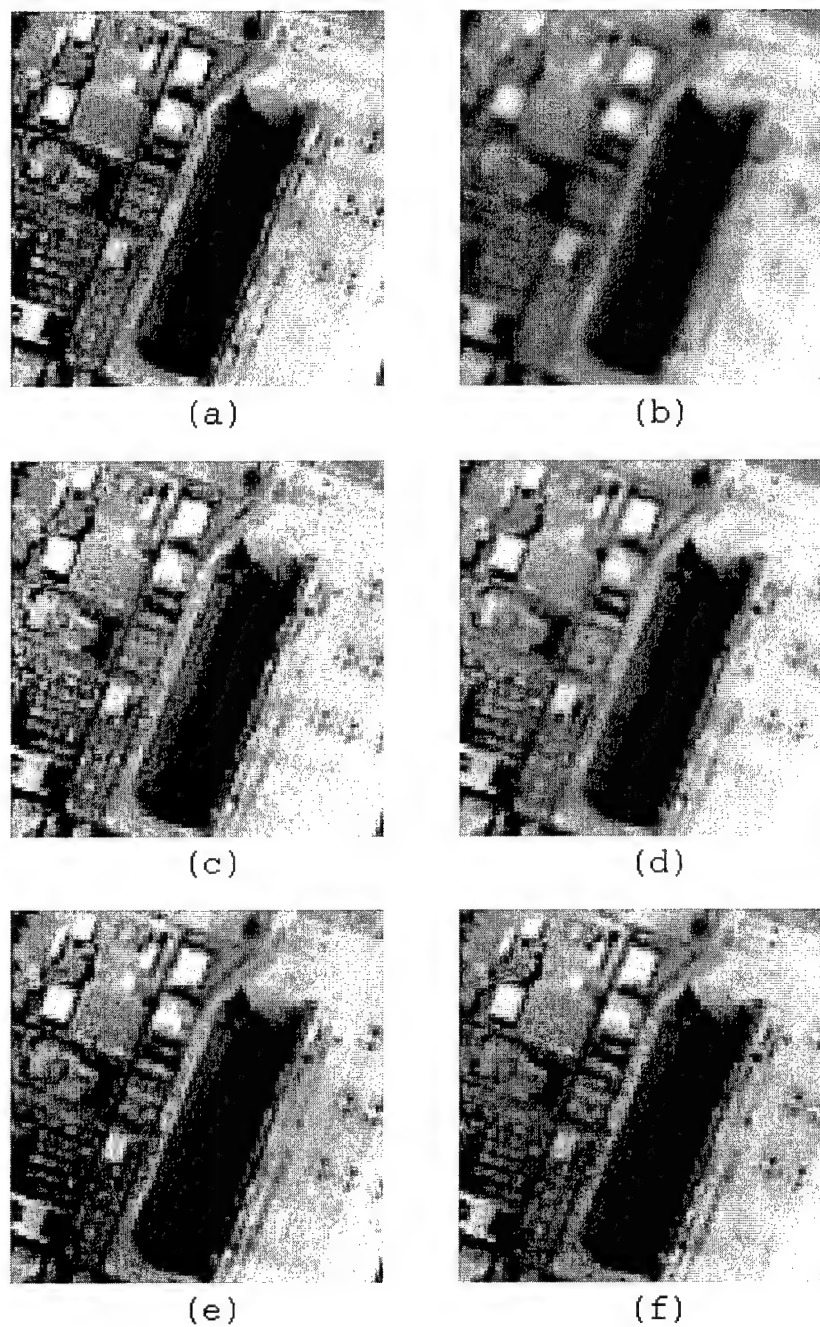


FIGURE 5.17. Restoration results. Enlarged section of image to show detail. (a) is original, (b) is degraded image (SNR 25.66), (c) results of fair WF (ISNR 2.85), (d) results using RS-DWT (ISNR 5.04), (e) results of the unfair WF (ISNR 5.82), (f) results using RS-DWT with wavelet domain WF (ISNR 5.04).

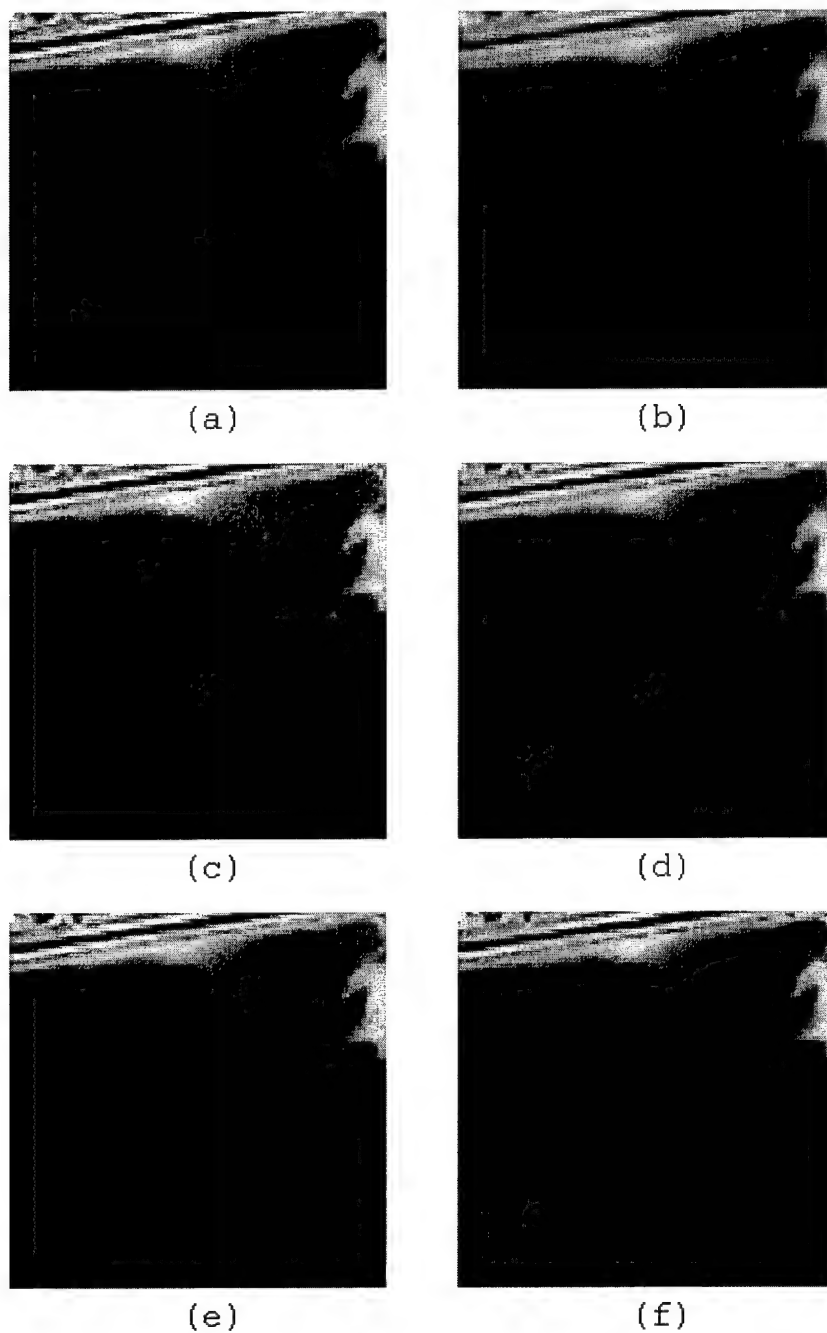


FIGURE 5.18. Restoration results. Enlarged section of image to show detail. (a) is original, (b) is degraded image (SNR 25.66), (c) results of fair WF (ISNR 2.85), (d) results using RS-DWT (ISNR 5.04), (e) results of the unfair WF (ISNR 5.82), (f) results using RS-DWT with wavelet domain WF (ISNR 5.04).

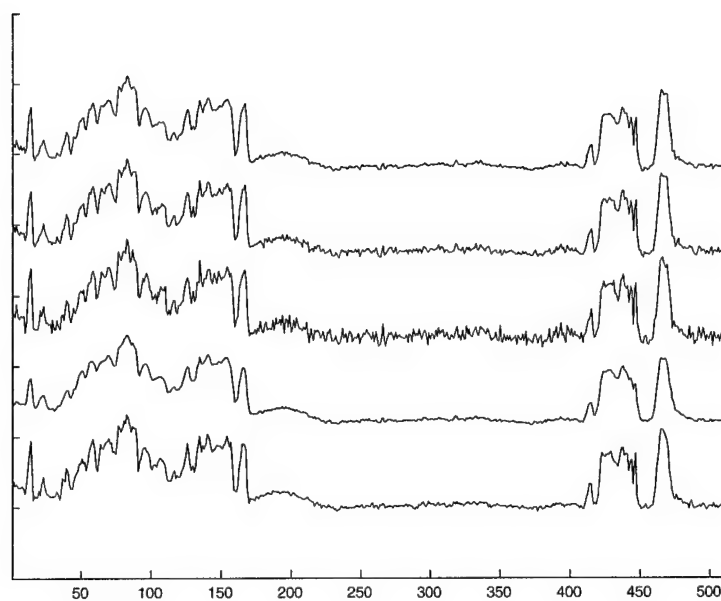


FIGURE 5.19. 1-D horizontal plots near the bottom (row 410 of 512) of the Urban image. The plots are offset from one another for visibility. From bottom to top, they are the original image, the degraded image, the fair WF estimate, the unfair WF estimate, and the RS-DWT estimate.

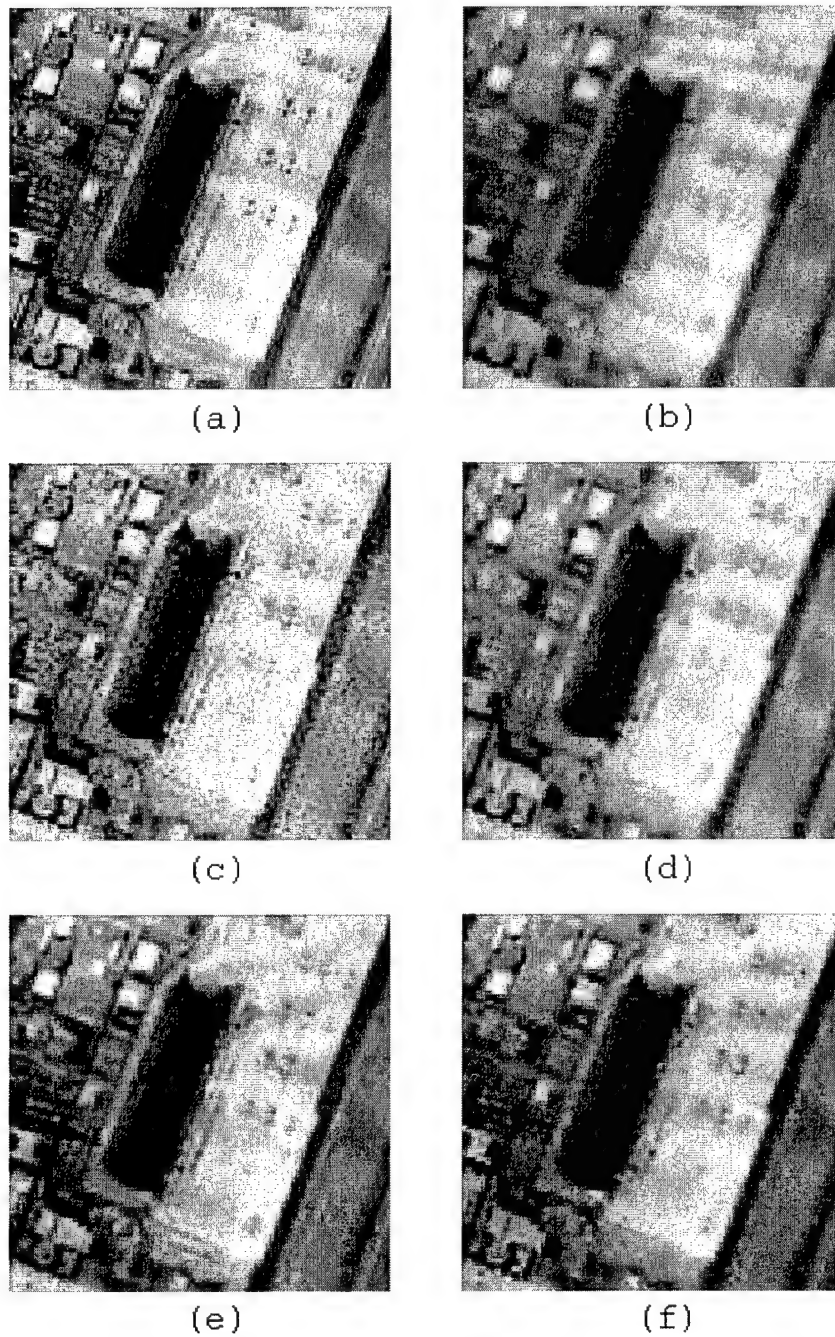


FIGURE 5.20. Restoration results. Enlarged section of image to show detail. (a) is original, (b) is degraded image (SNR 24.91), (c) results of fair WF (ISNR -0.36), (d) results using RS-DWT (ISNR 1.87), (e) results of the unfair WF (ISNR 2.63), (f) results using RS-DWT with wavelet domain WF (ISNR 2.06).

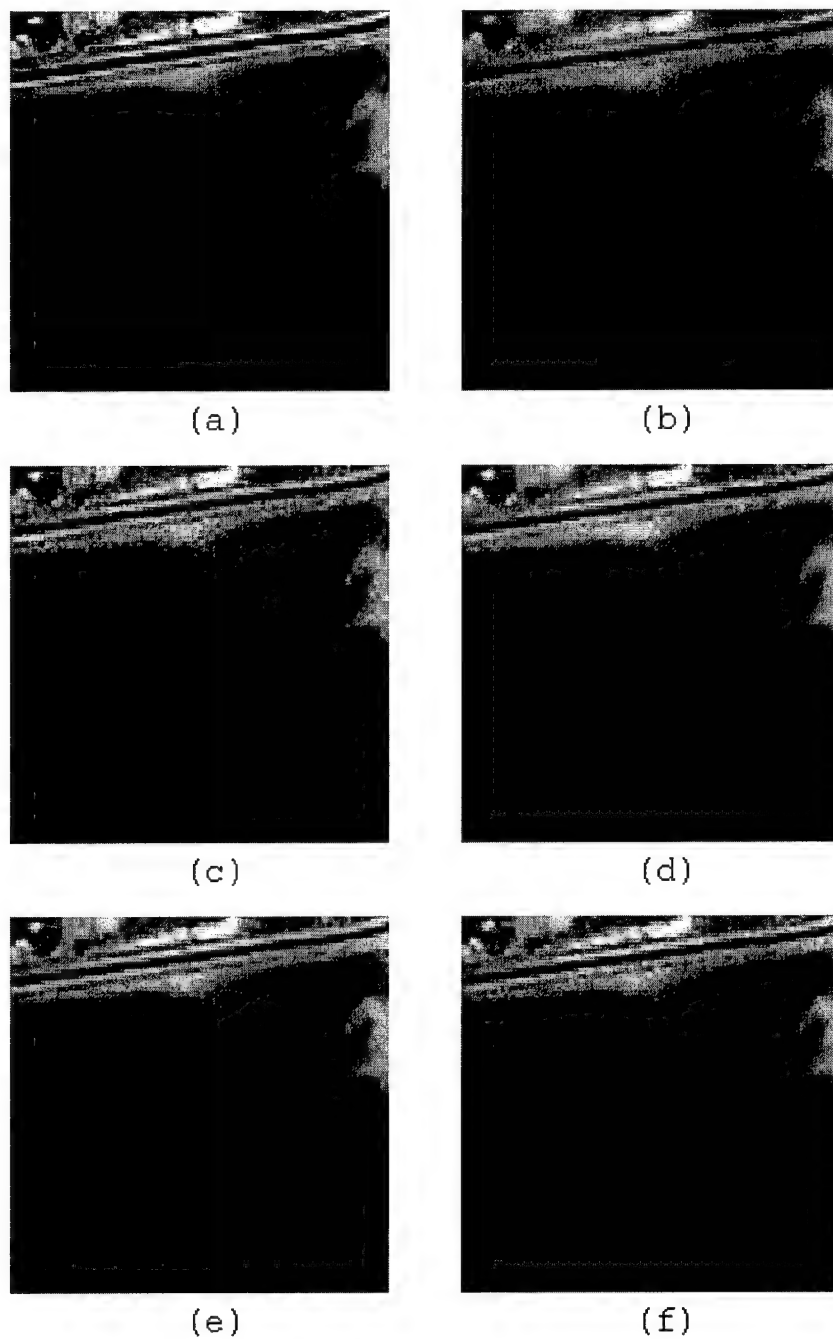


FIGURE 5.21. Restoration results. Enlarged section of image to show detail. (a) is original, (b) is degraded image (SNR 24.91), (c) results of fair WF (ISNR -0.36), (d) results using RS-DWT (ISNR 1.87), (e) results of the unfair WF (ISNR 2.63), (f) results using RS-DWT with wavelet domain WF (ISNR 2.06).

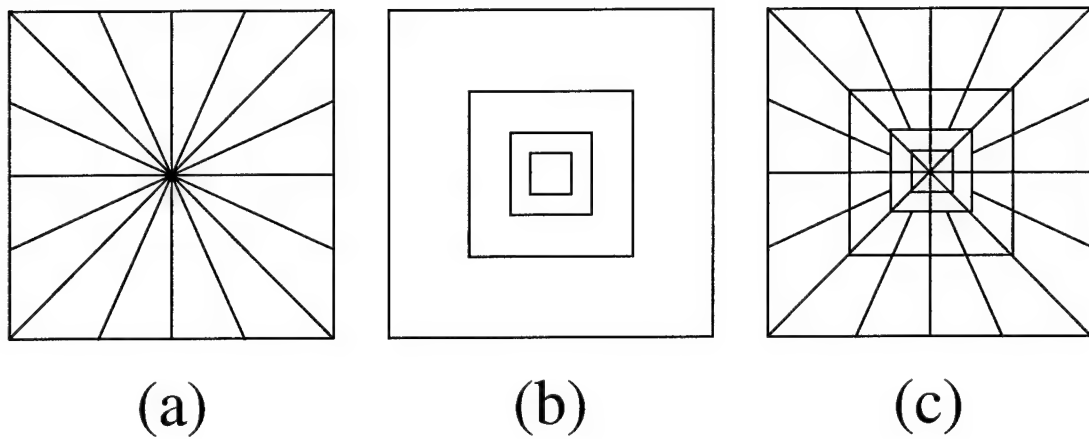


FIGURE 5.22. Frequency domain representations of (a) the directional filter bank, (b) a multiscale decomposition, (c) the PDFB with number of directions doubling at every other scale

CHAPTER 6

SUPER-RESOLUTION METHODS

6.1 Overview of Wavelets and Super-resolution

As implemented in the wavelet domain, super-resolution is fairly straightforward to describe: the determination of the next finest scale subband coefficients. I assume that the given image (which could be the result of deconvolution as discussed in the previous chapter) is an accurate representation of the scale J coefficients and the goal is to reconstruct an approximation of the scale $J + 1$ coefficients (\widehat{LH} , \widehat{HL} , and \widehat{HH}). Use of a single scale inverse DWT will yield a new image estimate, \hat{g} , as outlined in figure 6.1 for the case for a decimated separable transform.

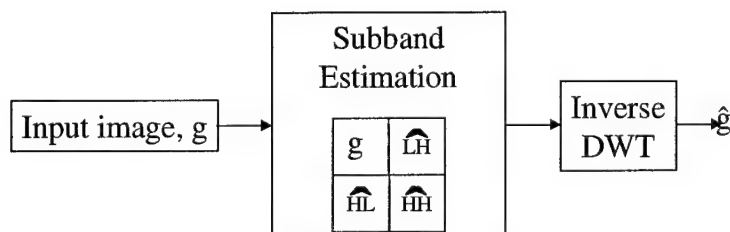


FIGURE 6.1. Overview of super-resolution process in the wavelet domain.

However, it is also possible to use an undecimated DWT or the RS-DWT. The three options are shown in figure 6.2. Use of a decimated DWT will require estimation of three $n \times n$ (decimated) scale J subbands. Use of a RS-DWT would require estimation of a single $2n \times 2n$ scale J subband while an undecimated, separable DWT would require estimation of three $2n \times 2n$ scale J subbands. The best choice is not clear a priori. Also note that the RS-DWT and undecimated DWT require the use of upsampled image data.

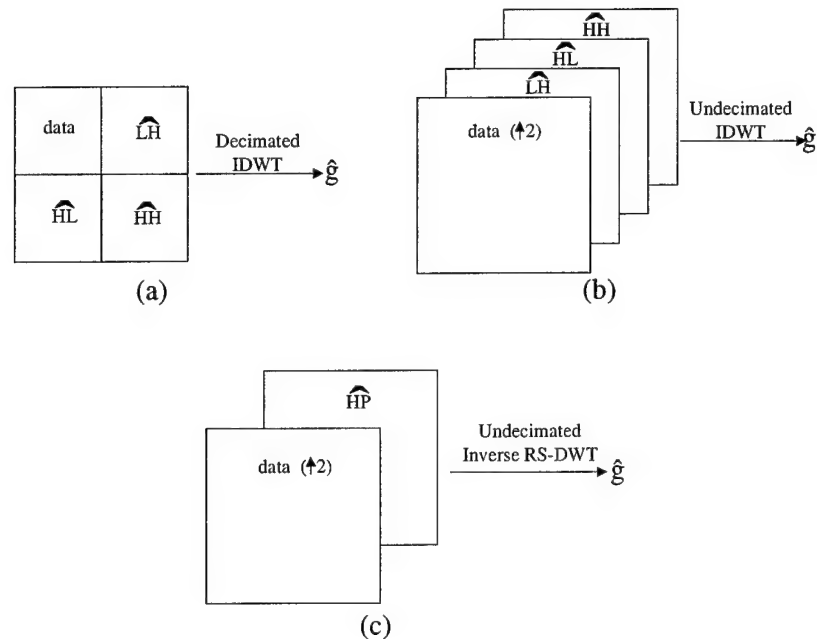


FIGURE 6.2. Three methods of wavelet based super-resolution: (a) uses the standard decimated DWT, (b) the undecimated transform, and (c) the RS-DWT.

As previously discussed, the two major aspects that are involved in successful super-resolution algorithms are *a priori knowledge* and *non-linear/spatially varying* techniques. Looking at *a priori knowledge* that may be useful for wavelet based super-resolution, positivity of the image is still a reasonable requirement. However, limited spatial extent is not of much use, a significant difference from previous techniques. This is due to the non-localized nature of the Fourier transform – spatial limitations in the image domain will cause expansion in the Fourier domain. However, the localized nature of the wavelet domain is inherently different and limited spatial extent implies zeros in the wavelet subbands, but this has only a local influence. One piece of *a priori* information is the known wavelet transform of the data. This can be used in predicting the estimated subbands in a statistical sense since we know DWTs generally maintain exponential decay across scales for edges and persistence of magnitude in general. Thus, the major piece of *a priori* information to take advantage of in the

wavelet domain is the nature of a DWT and the statistics of the known scales.

Looking at possible non-linear or spatially varying attributes that may be useful, we know that the decimated DWT is spatially varying. However, this is of little help in super-resolution as it is a result downsampling which decreases the frequency content, rather than the desired increase. In fact, the spatially invariant undecimated DWT has several advantages, notably the redundancy it supplies. The addition of new subband data beyond the original image is a spatially varying technique, and the main methodology used to achieve super-resolution.

As a first step towards wavelet based super-resolution, the next section discusses several initial experiments that were performed to quantify the possibility of super-resolution in the wavelet domain and to discuss the impact of the wavelet properties of sparsity, exponential decay, and persistence. This also includes discussion on possibility of using these for super-resolving imagery. The last section discusses the use of Vector Quantization to achieve super-resolution.

6.2 Feasibility Experiments

As a first step, I will discuss some simple experiments performed to determine the feasibility of achieving ‘meaningful’ bandwidth extrapolation of imagery. These are meant to look at the wavelet properties and how to use them in super-resolution algorithms. As basic set-up, I have used the standard images, Lenna, Urban, and Mandrill as shown in figure 6.3. In the simulation, I perform a band-limited down-sampling of the original image and use this as the input. The band-limiting is assumed to be an ideal low-pass filter in the frequency domain unless otherwise noted. In the experiments discussed here, I have used some knowledge of the original image in the restoration - i.e. assuming some knowledge that would be lost due to frequency band-limits in a real-world case. This is to simplify the experiments and determine the upper-bound on expected performance.



FIGURE 6.3. Images used in these experiments

As a measure of super-resolution performance, I will use spectral correlation coefficient (CC) images discussed in section 2.4.

In order to super-resolve, the estimated wavelet coefficient location and amplitude must be determined. First I will demonstrate the impact of sparsity - that knowing only a few coefficients provides significant super-resolution performance. Next I investigate the ability to predict the magnitude of the estimated coefficients. Then, I'll look at determining the location of these significant coefficients from known transform data.

6.2.1 Sparsity

The first set of experiments demonstrate the ability of wavelet based processing to perform significant super-resolution given imperfect knowledge of only very few of the estimated wavelet coefficients. This can be expected based on the sparsity property of the wavelet transform - since the energy is concentrated in few coefficients, knowledge of only those few coefficients can restore a significant amount of the information content.

For the experiments discussed below, I will downsample the original image by first applying a perfect low-pass filter in the frequency domain and then downsample by a factor of 2. I will then attempt to upsample this image back to the original size using 3 different techniques. First is Sinc Interpolation, which is upsampling via zero

insertion and applying a perfect low-pass filter. The second is wavelet interpolation: adding additional subbands with values of zero and then inverse transforming. The third is a wavelet algorithm that will add the additional subbands, but fill them with coefficients based on some knowledge of the true coefficients. For now, the assumption of perfect knowledge of location and sign will be made, but the magnitude will be varied. In the plots below, the magnitude will either be dithered from the true value by a random perturbation (AWGN, $\sigma = 20\%$ of magnitude) or simply use a single value for all of them, which was chosen as the mean of all significant coefficients. Additionally, I will calculate the frequency correlation coefficient image of the image estimate with the original as a measure of the meaningfulness of the interpolation. I will also show the average for all frequencies beyond the assumed system cut-off. Unless otherwise noted, the biorthogonal 9/7 wavelet system is used.

Figure 6.4 shows the results for the Lenna image, where 10% of the coefficients are retained. On the far left is a magnified portion of the original image to increase visibility of results. The next two columns are the interpolated images and the correlation images from the three techniques discussed above, with the mean of the CC image outside the original bandpass noted in the image title. As is obvious from examining the wavelet algorithm image and correlation image, significant ‘super-resolution’ has occurred. Note that neither the Sinc or wavelet interpolation yield any meaningful frequency content above the band-limit (the horizontal and vertical lines in the interpolated correlation images are due to edge effects). The box in each of the CC images denotes the band-limit of the downsampled image. Figure 6.5 shows the results for only 1% of the coefficients in the new subbands. While performance does decrease, there is still significant meaningful frequency content above the cut-off.

To further understand performance over a variety of cases, the average correlation coefficient (outside the band-limit) can be plotted vs. the percentage of coefficients included. This is shown in figure 6.6 for three cases of what magnitude was used: actual magnitudes (solid), random perturbation (dashed), and mean value (dotted).

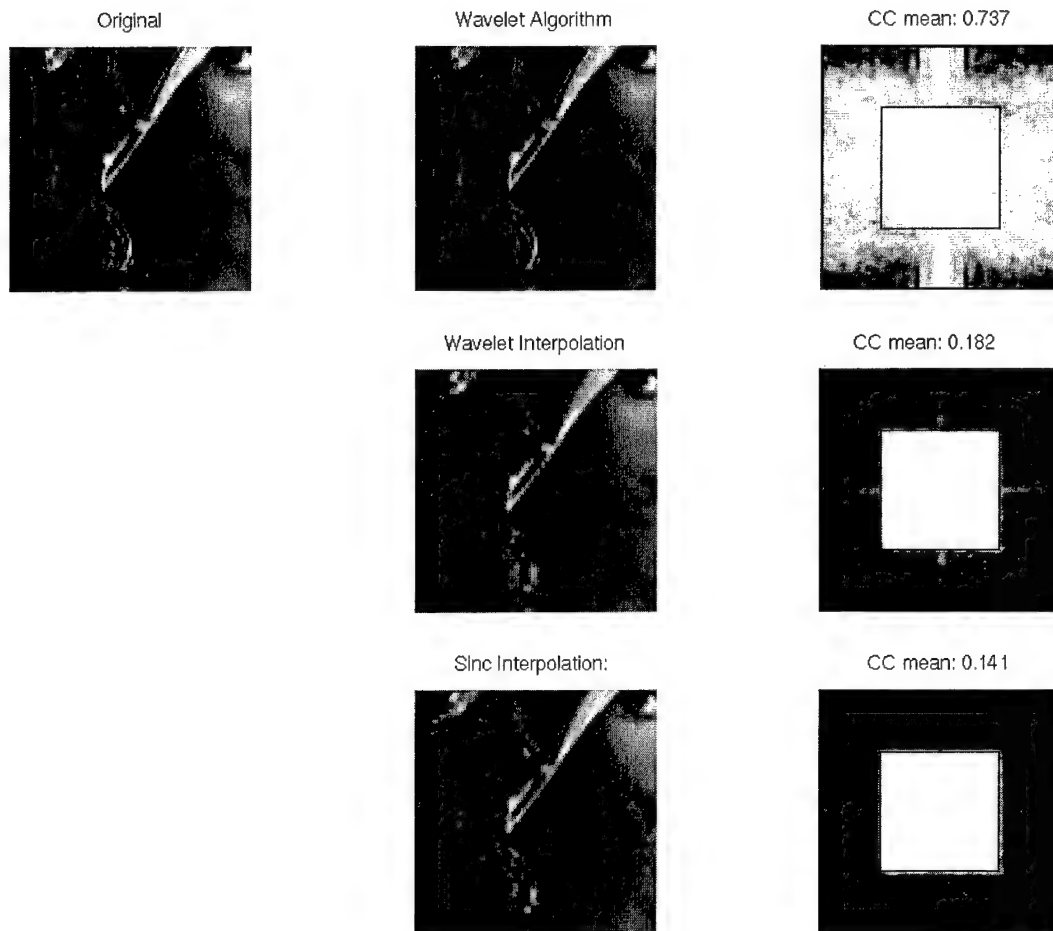


FIGURE 6.4. Interpolation results for 10% of coefficients and random perturbations on the wavelet coefficients.

For all three cases, the correlation average rises fairly quickly as coefficient percentage is increased, and then evens out. For very low percentages, the results are similar, but eventually the actual magnitudes wins out (as expected), but the random perturbation on the magnitudes does not have a large effect. The mean-value case approaches a maximum around 9-10% and then drops due to the fact that as more coefficients are averaged, the mean is decreasing in value. Almost identical results are also achieved using the RS-DWT discussed in the previous chapter. Figure 6.7 shows the results for each of the three images with random perturbation on the coefficient magnitudes.

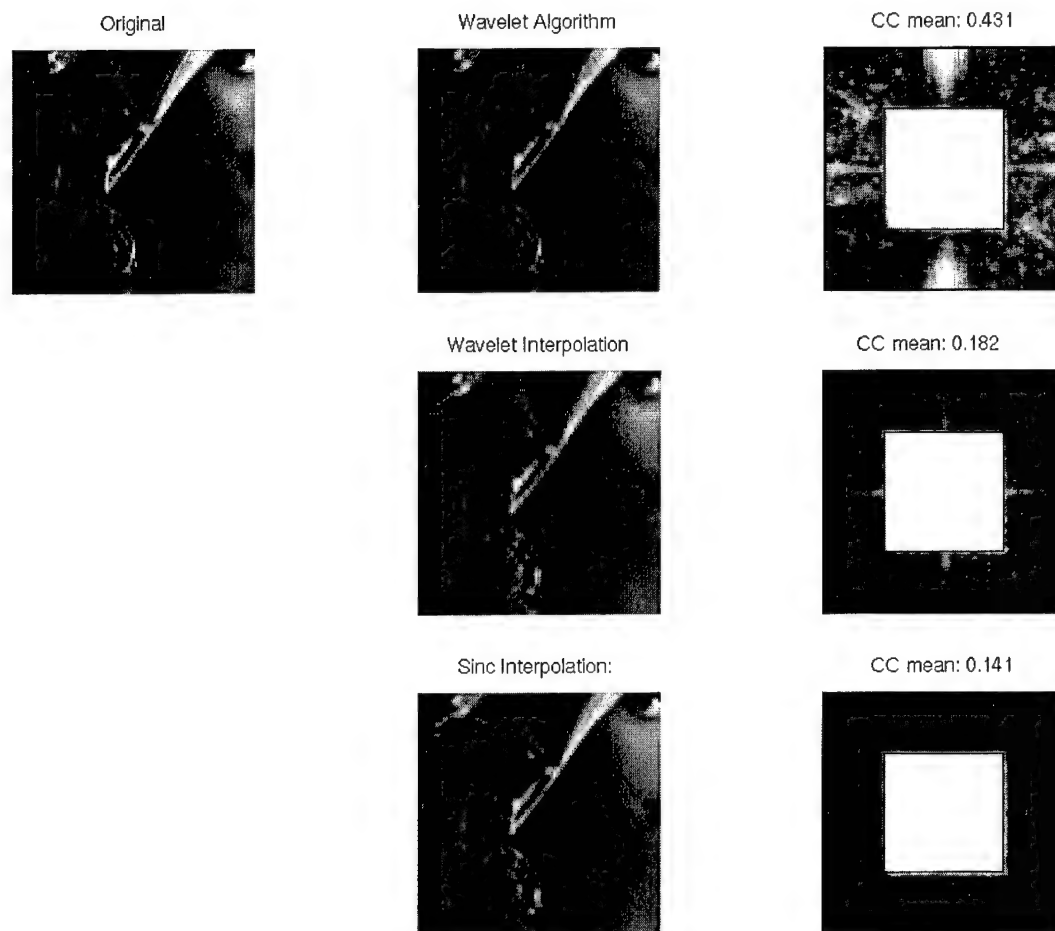


FIGURE 6.5. Interpolation results for 1% of the coefficients and random perturbations on the wavelet coefficients.

The same general form is seen in all three cases. One note is that the correlation does not go to zero. In fact, even the correlation of two independent white noise images will not be zero, but will depend on the correlation neighborhood size (k in equation 2.30) used. For $k = 3$, the default used, the value will be about 0.128.

The primary conclusion from these experiments is that a significant amount of super-resolution is possible from knowledge of only a few coefficients.

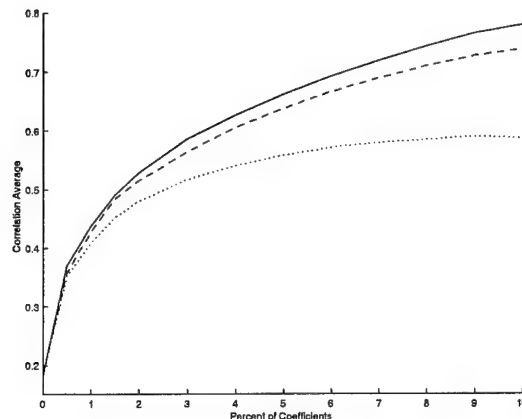


FIGURE 6.6. Plot of average correlation coefficient vs. percentage of coefficients included for the use of actual magnitudes (solid), random perturbation (dashed), and mean value (dotted).

6.2.2 Magnitude Extrapolation

The next set of experiments evaluated the ability to predict the mean value of the extrapolated subband based on data from the downsampled image DWT. As discussed above, there is a nearly exponential decay of the coefficients across scales for images dominated by edges. In practice, this holds true for the large coefficients in most natural imagery as they tend to represent edge content. Figure 6.8 shows the result of predicting the mean of the 1% most significant coefficients from the mean of the 1% most significant coefficients in the subbands of the downsampled imagery. From left to right, the columns are the HL, HH, and then LH bands. From top to bottom, they are the results from Lenna, Urban, and then Mandrill. Note that the plots show the linear fit line (semilog scale is used, thus exponential relationship is linear) along with the downsampled means (circles) of the known data and the true value (x's) of the to be estimated subbands. The title on each shows the percentage error between prediction and actual values. The next figure is the results, shown in the same format as the previous section, when the predicted magnitude is used for all 1% of the most significant coefficients. The location and sign of the coefficients was

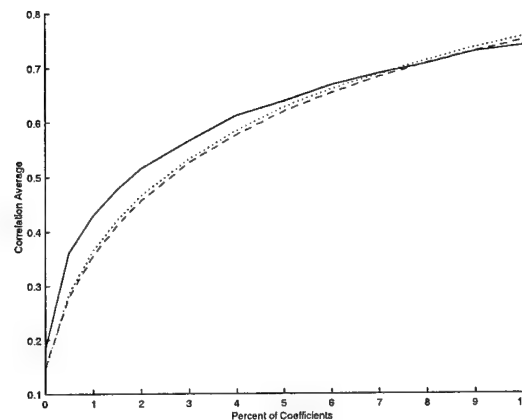


FIGURE 6.7. Correlation averages for Lenna (solid), Urban (dotted) and Mandrill (dashed) given random perturbation of the magnitudes.

assumed known. Note that even this imperfect knowledge of magnitude still yields meaningful bandwidth extrapolation.

6.2.3 Determination of Significance

The last major experiment in the feasibility of wavelet based super-resolution was looking at the ability to know which of the extrapolated subband coefficients are significant. The input we have in this determination is the significance of the lower resolution subbands. It is reasonable to expect some correlation in this significance, especially given the persistence property of the wavelet transform. To correlate the magnitudes across scales a further discussion on parent/child relationships is necessary.

In the typical, decimated wavelet transform, the meaning of parent and child is well defined, with each parent having four children for a 2-D image, as shown in figure 6.10. However, as discussed previously, the undecimated transform yields benefits stemming from shift invariance and redundancy. Since each of the subbands in an undecimated DWT is the same size, it seems reasonable that each coefficient should be assigned a single child – the coefficient located in the identical position and the

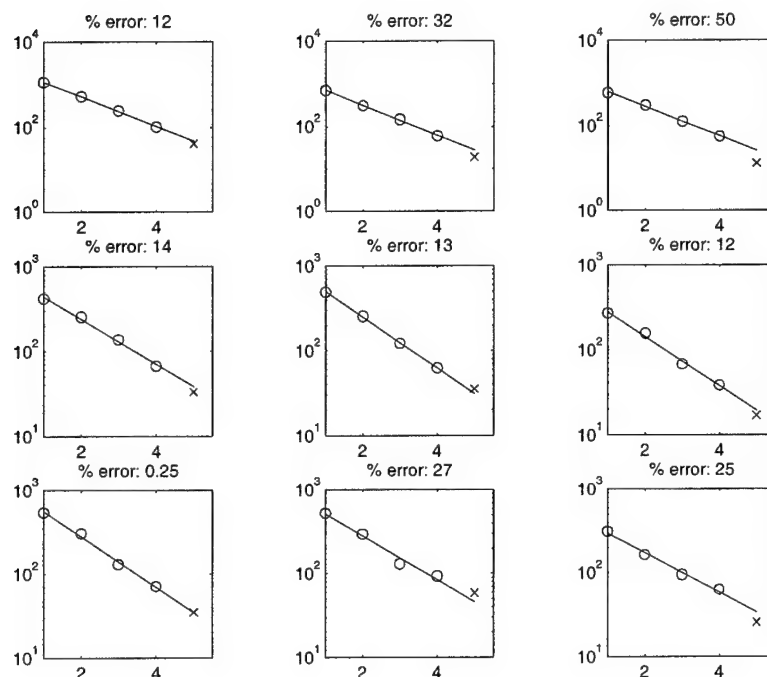


FIGURE 6.8. Results of predicting new subband magnitude from exponential decay. Left to right are LH, HL, HH subbands. Top to bottom are Lena, Urban, and Mandrill images.

next finer scale. While accurate to assign the more detailed coefficients as children, it is not a complete view of the children of any coefficient.

To understand the relationships in an undecimated transform, we need to look more closely at how the coefficients are calculated. In order to simplify the discussion, a 1-D example will be shown, based on the Haar wavelet (simple difference of adjoining coefficients are the wavelet coefficients). The results, however are identical for any finite wavelet and are extendable to 2-D in the typical separable manner – i.e. a coefficient has children in the horizontal and vertical directions. Figure 6.11(a) shows a single scale decomposition. The next scale (coarser) decimated wavelet coefficients are calculated based on two adjoining coefficients. In 6.11(b), we extend this to the undecimated transform. We see the results are the same as the decimated case (solid lines) with additional coefficients interspersed in every other position (dotted lines).

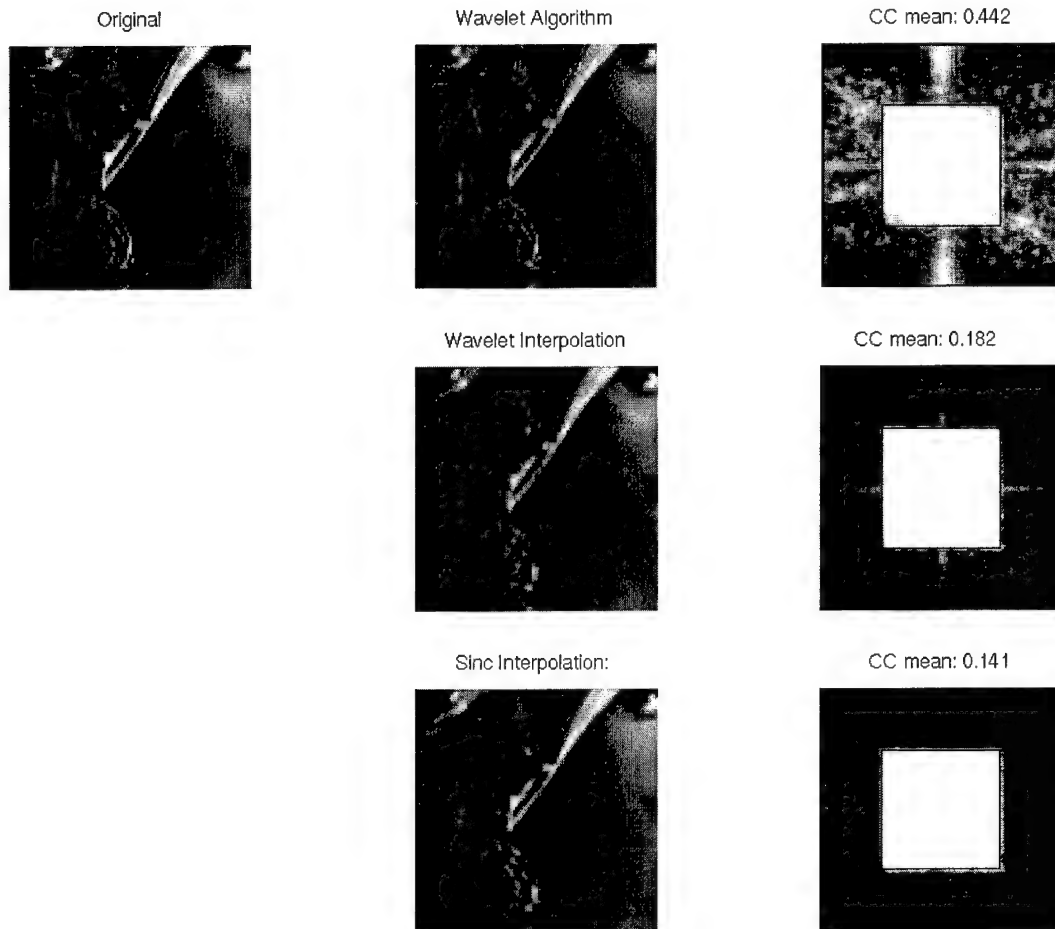


FIGURE 6.9. Results of wavelet super-resolution algorithm using known location, but predicted magnitude of 1% of the coefficients.

However, the parentage is now not as clear: the ' b ' coefficient derives not only from ' $a - b$,' but also ' $b - c$.' This is understood from viewing the undecimated transform as a compilation of the decimated transform for all applicable shifts (section 3.5). In 6.11(b), the solid and dashed lines are the results for no shift and a single pixel shift. Obviously, this is the only applicable shift, as a shift of 2 pixels will only replicate the zero-shift result (albeit shifted by one pixel). Thus, each pixel has two parents, just as each parent has two children.

This result can then be extended to a two-scale decomposition as shown in figure

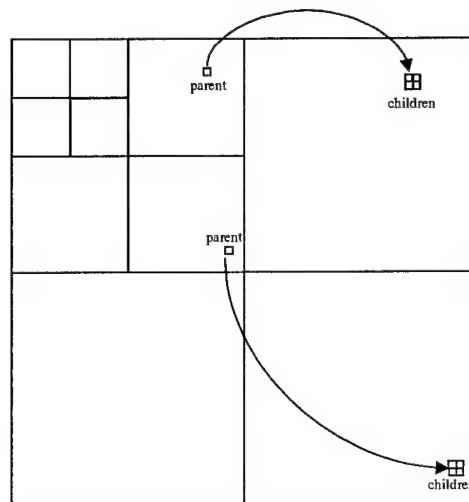


FIGURE 6.10. Parent-child relationship in a 2-D DWT

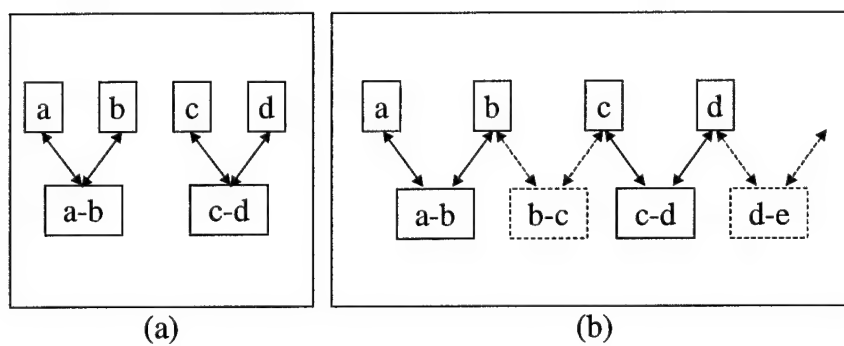


FIGURE 6.11. Parent-Child relationships for a 1-D signal for (a) Decimated and (b) Undecimated DWTs.

6.12. At the first level of decomposition, we have the results as above, and at the second level, we again calculate differences. However, since the filter must be upsampled, it is no longer differences of adjoining coefficients. Thus the first undecimated coefficient ($(a - b) - (c - d)$) is parent to the first and third at the next finest scale ($a - b$ and $c - d$). It also has four grandchildren: the first four coefficients at the finest scale. Each finest scale coefficient will also have four grandparents. Thus there is a n -to- n relationship between the scales, but not necessarily always between adjoining coefficients.

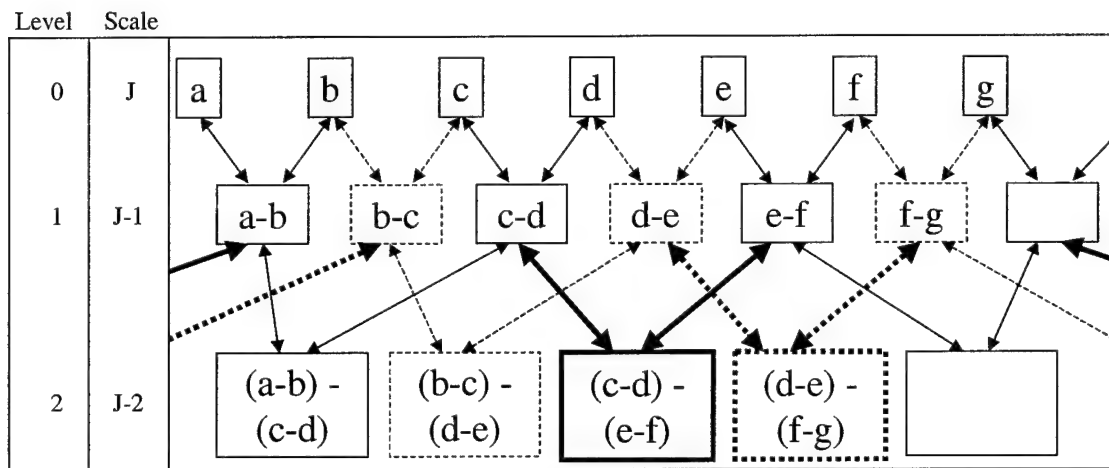


FIGURE 6.12. Parent-Child relationships for a 1-D signal for an Undecimated DWT.

The key result is that while there is a parent-child relationship for pixels in identical locations, this is only part of the relationships that exist in the undecimated form. In order to follow the ‘family’ relationships, the above map must be used.

To quantify the ability to predict the significant coefficients in the next scale, I will use the following two definitions which are based on the parent/child relationship.

- Successful spawning rate (SSR): The percentage of significant coefficients at one scale that have at least one of their children being significant.

- Legitimacy rate (LR): The percentage of significant children with significant parents.

These are in some sense, complements of each other in that SSR measures the forward dependence of significance (is significance passed on), while LR measures the backward dependence (was significance inherited). A large SSR implies that the majority of significant coefficients pass on significance. A large LR implies that most of the significant coefficients will be found by searching only the children of significant parents. Obviously, we would like both to be high.

While the definitions appear similar, there is an important distinction. As an example, we can achieve a SSR of 100% if we use a threshold such that only a single coefficient is significant and it has a significant child. While achieving this perfect SSR, we will only restore 1 significant child. Thus, LR is complementary in the sense that it measures how many coefficients we can find by looking only at children of significant parents. In the above example, we would only find a single significant location – a very low LR. Remember that for super-resolution we want to use the data from scales up to J to predict a new $J + 1^{st}$ scale. These quantities measure our ability to predict the locations of the most significant of the $J + 1^{st}$ scale coefficients.

It is possible to set the significance levels at different levels for different scales, e.g. we can use a 10% threshold at scale J , and a 5% threshold at scale $J + 1$. Intuitively, this will increase the LR, since we have more possible parents for the children, however, it should decrease the SSR since we are adding more parents without adding more children. This can lead to an optimization problem: At what level do we set the thresholds for scale J such that the SSR and LR are optimal for predicting the locations of a given percentage of the most significant coefficients at scale $J + 1$?

The results discussed below are based on calculating the DWT of a given image, determining binary significance maps based on the (scale dependent) threshold, and then calculating SSR and LR for prediction of the (known) highest scale significance

data from the next highest scale data. In addition to the same 3 images used before, I have also used a synthetic test image show in figure 6.13. The reason for this is that wavelets (meeting certain smoothness criteria) will yield a DWT that decays exponentially across the scales for step edges. This regularity in the DWT coefficients, will be higher for the test image than real imagery due to the exact step edges. Thus, the test image should hint at an upper-bound in performance.

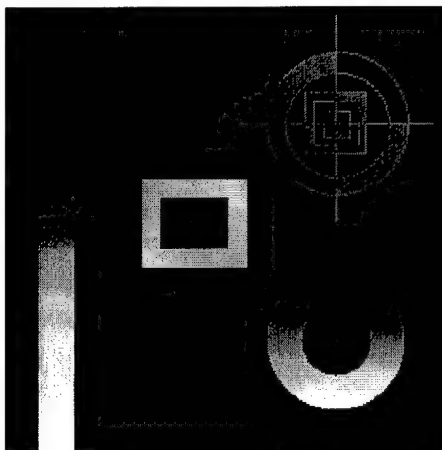


FIGURE 6.13. Test image used in experiments

The results are shown in figures 6.14 and 6.15. Both of these plot the calculated SSR and LR for the four images, using a significant percentage for the J^{th} scale (parents) of 10% significance for figure 6.14, and 5% in figure 6.15. The x-axis then shows the significance threshold (in terms of percent) of the $J + 1^{st}$ scale (children) that was used, varying from .01 to .30 (1% to 30%). As expected, the SSR increases from left to right. As the percentage of children coefficients deemed significant is increased, it is more likely that they will have significant parents. Likewise, the LR will decrease – as more children coefficients are denoted as significant, they are less likely to have significant parents, as the number of significant parents is held constant. Note the SSR and LR rates (y-axis) are the fraction of significant coefficients, not of the total number of coefficients in the subband.

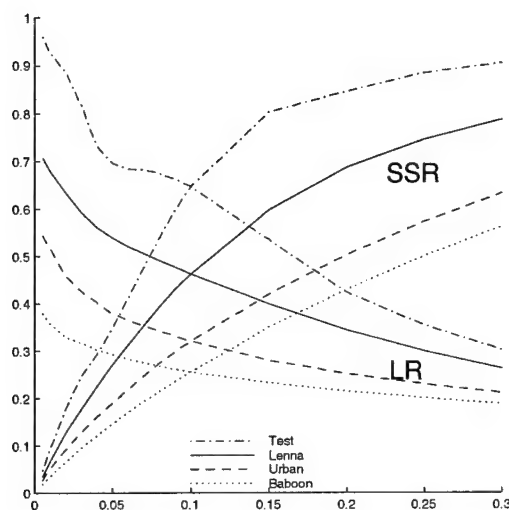


FIGURE 6.14. Plot of SSR and LR for 10% significance at scale j

It is interesting to note the result that the images can consistently be ranked in the order Test, Lenna, Urban, Mandrill, where the order denotes the relative levels of SSR and LR. Test gave the best results, with consistently higher SSR and LR, while Mandrill gave the lowest. This makes sense by looking at the images: Test, being a synthetic image with sharp edges and no texture/noise, should provide the strongest persistence across scales. Lenna is fairly regular with sharp edges, Urban a little less so, whereas Mandrill is composed primarily of what is termed texture, rather than edges. Texture is an important aspect to consider. As opposed to edges, which give exponential decay across scales, the relationship between scales for texture is dependent on the texture. How to deal with texture is something to be considered in super-resolution if images such as Mandrill will be used as input.

6.2.4 Discussion

In order to super-resolve in the wavelet domain, we need to know the location and magnitude of the wavelet coefficients. While it was shown that even with a rough guess at the magnitude from previous scales used in only 1% of the coefficients,

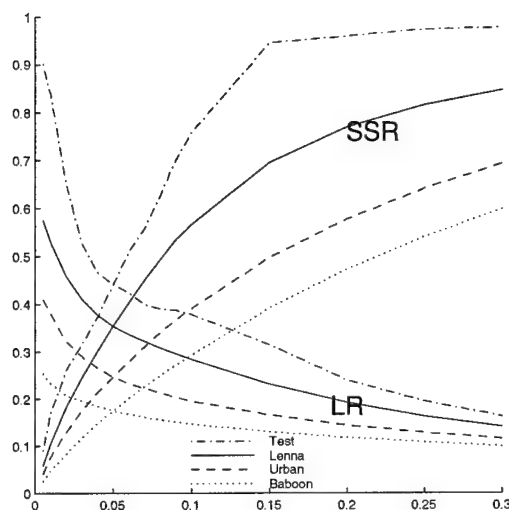


FIGURE 6.15. Plot of SSR and LR for 5% significance at scale j

super-resolution is feasible. However, not surprisingly, the determination of location is the most difficult part. Use of the significance in the previous scale is not a good prediction of significance in the next scale. In order to get decent performance, an alternative method of location determination must be used.

While these experiments show that modest super-resolution is possible without total knowledge of the subband coefficients, the determination of location is a critical flaw for any real super-resolution attempt. Using the techniques above, there is not enough information to adequately predict the location of the coefficients. An additional complication is the sign of the coefficients. The experiments above are based on calculating the magnitude from previous scale data. Even if the locations could be determined, the sign is still unknown. Studies as to the ability to predict the sign only [73] point out that the sign itself is very hard to predict.

This difficulty in super-resolution is not totally unexpected from the experience gained in image compression using wavelets. In compression, most of the bits are used coding the largest coefficients and few (if any) on the small coefficients, with minimal loss in the visual quality of the imagery. In the magnitude determina-

tion experiments, a small percentage of coefficients were added to the highest scale sub-band, with a quantized magnitude – very similar to image compression with 3 quantized values $(-\beta, 0, \beta)$. As in compression, decent results are possible with such an approach. However, in compression, the location is also coded, which is missing in super-resolution. The fact that only a few wavelet coefficients can maintain a decent visual quality of the image is well known throughout the compression community. However, this work shows that they also recreate a significant portion of the true object frequency spectrum in the band of interest. It would be an interesting experiment to look at the spectral correlation of compressed images from various algorithms and compare the results to see how well this measure correlates with subjective visual quality and if it is better than other measures such as SNR.

6.3 Vector Quantization

In order to capture the expected relationship between the known wavelet coefficients to the unknown coefficients, a form of vector quantization, non-linear interpolative vector quantization (NLIVQ) will be used. VQ has been used successfully in numerous image compression techniques[74], and more recently for image processing and restoration[12, 75, 76]. For the purposes of this work, NLIVQ will be used in a pattern recognition sense - capturing patterns between wavelet coefficients in the degraded and true images.

6.3.1 Overview

Given random processes Q_x and Q_y with ranges of X and Y , the NLIVQ process maps a vector in X to the respective (quantized) vector in Y ,

$$V : X \rightarrow Y \tag{6.1}$$

such that the average distortion is minimized. Often, the approach is to optimize the VQ for Q_x and then design a VQ for Q_y that is conditionally optimal. When the exact probability distributions are not known, the optimization can be approximated using a large set of representative training data. In this application, the Q_x and Q_y are blurred and original wavelet coefficient vectors. A VQ for Q_x is designed based on the expected probability distribution for wavelet coefficients. The VQ for Q_y is derived from the training data as the centroid of all vectors in Q_y (original data) that have the same codebook index from Q_x (blurred data).

This process uses the following steps:

Encoder design

1. For each of the training images, the undecimated DWT is calculated. At each pixel location, a vector is formed from the wavelet coefficients. The vector can be formed many ways, such as from a single coefficient from each scale, or the neighboring coefficients.
2. Given an encoding rate, the bits are allocated to the individual vector coefficients so as to minimize the MSE distortion.
3. Using a Laplacian distribution as an estimation of the probability distribution of wavelet coefficients, the coefficients are scalar quantized using the number of bits from the previous step. The encoder index is the binary concatenation of the scalar quantized values.

Decoder design

1. For each pixel location, calculate the encoder index, i , from the blurred data. Add the vector derived from the original data at the same pixel location to a variable, $D(i)$, and track the number of such vectors for each index i .

2. After all the training data is exhausted, compute the average of each of the $D(i)$

Image restoration

1. For each pixel location in the blurred image calculate the encoder index, i , as above
2. Replace the current vector with $D(i)$.

Sheppard[12] was the first to introduce NLIVQ for image restoration where the vectors were defined from the DCT coefficients of a 3 x 3 pixel block. Improved performance was achieved using a ‘lapped’ technique[76], where the 3 x 3 pixel blocks were overlapped and averaged. However, this is at the cost of compression since now it is required to store a codeword for each pixel rather than each 3 x 3 block. In [75], the use of wavelets is proposed instead of the DCT to remove the blocking artifacts while retaining the ability to jointly compress and restore the image.

The above algorithm calculates the wavelet coefficients only, which ignores the scaling coefficients. In this work, I will follow the concept in [12] and calculate the scaling coefficient values by using a Wiener filter on the blurred image, transforming into the wavelet domain, and then extracting only the scaling coefficient data. Since the scaling coefficients include only low frequency information, it is assumed a Wiener filter restores the information well enough. In practice, the difference between results using this method and the true data is small, as was seen in [12].

6.3.2 Experimental Results

Using NLIVQ as described above, the goal is to super-resolve imagery. Thus, there will not be an attempt to compress the imagery. In order to provide redundancy in the representation, I will use undecimated transforms. Due to the redundancy in

an undecimated transform, there is actually an increase in memory to represent the image. In the results below, a set of 53 urban images, each of size 512 x 512, were used. At least one of the images was left out of the training data and used in the restoration attempts.

When only band-pass restoration was attempted (incoherent OTF blurring operator) using a separable DWT, an ISNR over 3 dB was noted as shown in figure 6.16 which contains only a portion of the entire image so that details are visible. While a demonstration of NLIVQ, the results are below expected performance from other typical algorithms. For these, the highest 3 subbands were used to form the vector. When combined with the three directions (LH, HL, HH), this yields a vector length of 9, and 15 bits were used. Note in figure 6.16(d) which shows the spectral correlation from which we can get an understanding of where the improvement comes. Note that at the higher frequencies, there is less correlation. Eliminating the highest resolution subband ($\frac{\pi}{2} < \omega_x, \omega_y < \pi$) in the restoration lowered the ISNR to 2.15 dB. Thus, there is still meaningful information here but much of the improvement seems to stem from the lower frequency content.

Next, the RS-DWT basis was used for band-pass restoration. A full packet transform of the RS-DWT was used which yielded a uniform distribution as shown graphically in figure 6.17. As opposed to the separable transform which has each of the three directions (HL, LH, HH), the RS-DWT has only a single subband at each level. Thus, a three-level full-packet transform yields 7 wavelet subbands for the RS-DWT versus 63 for the separable DWT. This greatly simplifies matters, allowing 3-level (rather than 2-level) transforms while maintaining a reasonable rate for the VQ. The results are shown in figure 6.18. Note the improvement in performance from the separable case, the ISNR nearly doubles. I believe this improvement is mainly due to the increased rate that can be achieved per scale when using the RS-DWT which does not have directional subbands (HL, LH, HH).

When the diffraction cut-off is placed below the folding frequency, super-resolution

is possible. Otherwise, critically sampled data can be upsampled to allow for the possibility of super-resolution. Often, the cut-off is set at half the folding frequency, but I will discuss other cases below. The wavelet coefficients in any subband above the cut-off frequency represent the super-resolved data. Note that as discussed before, under-sampled data is not discussed in this research as it requires a different definition of super-resolution and different algorithms.

First, separable transform were used. For this case, 4 resolution levels were used, yielding a vector length of 12 while using a 20 bit quantizer. Since no bits are assigned to the highest subband, the concept is the same as pass-band restoration using separable DWTs above, adding in the estimation of the next higher resolution subband. The results, figure 6.19, show the same ISNR in the band-pass only case, although almost double the results from [75]. This is probably due to the extra coefficients included in my vector below the diffraction cut-off, which is not feasible when a uniform DWT packet basis is used. While there is frequency content in the super-resolution band, it does not appear to be meaningful (associated with the true object frequency data), as can be seen in figure 6.19(d).

While [75] shows some frequency content beyond the OTF cut-off, the overall improvement in the imagery was modest (no greater than 1.75 dB), and it is difficult to assess whether this is meaningful extension of the frequency content. While they use of a Wiener filter to improve the blurred image prior to the NLIVQ technique, it is difficult to assess whether the improvement stems from further band-pass improvement or true super-resolution.

Since the predominant interest in this work is in super-resolution and band-pass restoration complicates the determination of this, the remainder of the experiments assume perfect band-pass knowledge. Thus, any improvement in comparison to the original image will be due to super-resolution. While not necessarily feasible to expect perfect band-pass knowledge, this functions as a reasonable limiting case.

I will also introduce a new metric, the SR-ISNR, as a measure of the super-

resolution performance, separating out effects in the band-pass region.

$$SR-ISNR = SNR(truth, restored) - SNR(truth, band-pass restored) \quad (6.2a)$$

$$= 10 \log_{10} \left(\frac{\langle |f - \hat{f}_{band-pass}|^2 \rangle}{\langle |f - \hat{f}|^2 \rangle} \right) \quad (6.2b)$$

where f is the true object, \hat{f} is the NLIVQ restoration, and $\hat{f}_{band-pass}$ is the band-pass portion of the NLIVQ restoration – excluding any super-resolved frequency content.

Using the same series of 53 urban images, the object frequency will be band-limited (ideal low-pass filter, circular symmetry) below the folding frequency. These images will be used as the input to the restoration process – hoping to restore the lost frequency information. Figure 6.20 shows the results for a diffraction cut-off at $\frac{3}{4}$ the folding frequency. This is used rather than $\frac{1}{2}$ to allow the NLIVQ more subbands of known data from which to extrapolate. While the results show added frequency content, they do not show any significant signs of super-resolution. The spectral correlation plot may show some correlations (the mean in the super-resolved frequencies is 25% higher, 0.17 versus 0.13), but it is still quite small. Note the the ISNR is negative – the NLIVQ actually decreased the SNR relative to truth data. This is due to the quantization that takes place in the NLIVQ process – quantization for those wavelet coefficients below the diffraction cut-off will reduce the SNR relative to truth. The SR-ISNR is 0.15 dB and thus there is some super-resolution, albeit quite small.

These results are consistent with a range of experiments conducted to estimate the super-resolved subbands via NLIVQ. Multiple experiments were conducted to look at vectors to use in the NLIVQ process, rather than simply the same pixel in each wavelet subband. Using a neighborhood of pixels in the parent subband and other techniques did not yield any significant super-resolution, with SR-ISNR's all below 0.2 dB. The cases when the image was part of the training data, did show greater

improvement, up to 4-5 dB in SR-ISNR, but this is not a reasonable expectation in practical situations.

A visual examination of the true and estimated subbands shows that the process does somewhat replicate the form seen in the true data, as in figure 6.21. But as seen by more detailed examination and in the 1-D plot of the respective coefficients (figure 6.22), the NLIVQ estimate does not capture all the detail in the true data – it appears much of the detail in the true wavelet coefficients is lost in the NLIVQ estimate.

It is interesting to revisit the original concept of using DCT based NLIVQ for image restoration. In general, these results based on DCTs were superior to the wavelet based results above. The results for a DCT based NLIVQ estimation (using the lapped approach) is shown in figure 6.23 for perfect band-pass knowledge out to half the folding frequency. Figure 6.25 shows the complex correlation plot and 6.24 shows a log-compressed plot of the frequency content. In this case, the SR-ISNR was 0.95 dB, significantly higher than in the wavelet case.

These results point out the difficulty of taking advantage of correlations in the wavelet domain to produce super-resolution. The DCT actually provides a better transform, probably due to the improved decorrelation that occurs in the wavelet transform while the DCT leaves more correlation between coefficients. An interesting approach for future work would be to investigate other multi-resolution transforms, such as the Laplacian Pyramid, for NLIVQ based super-resolution performance, specifically when the filters are designed so as to leave correlations in the data.

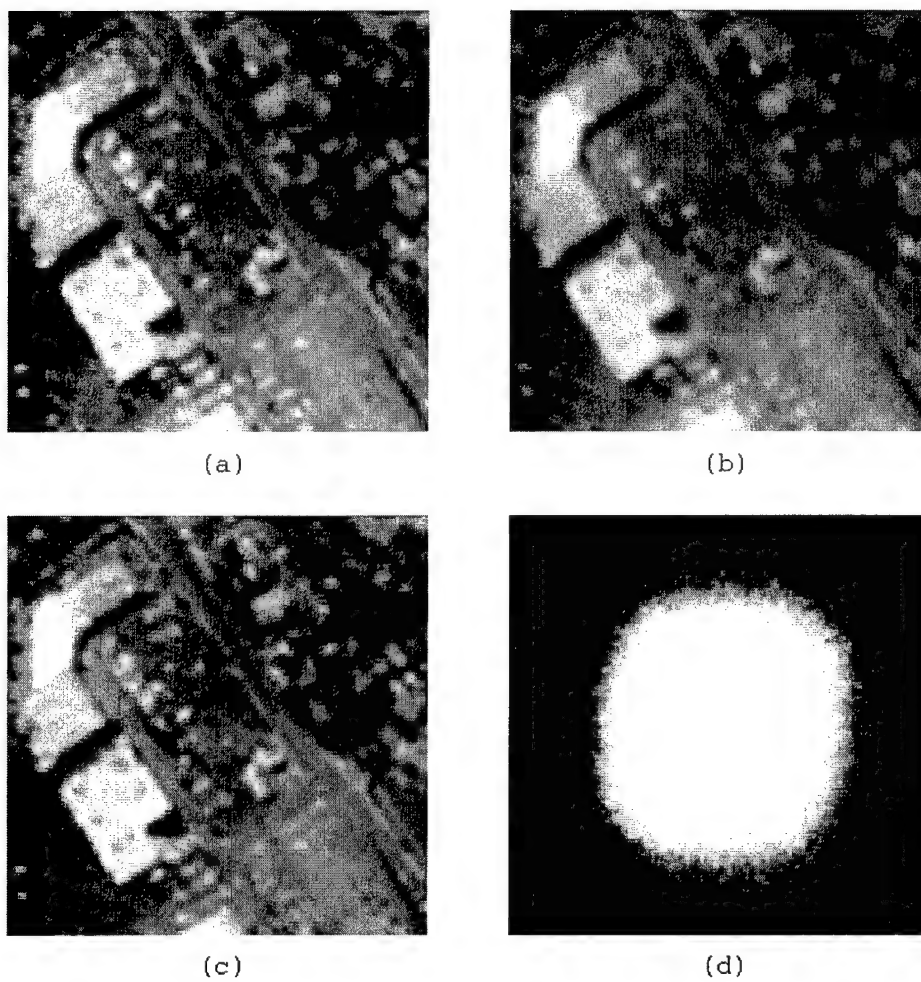
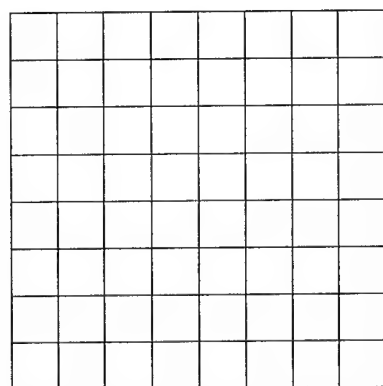
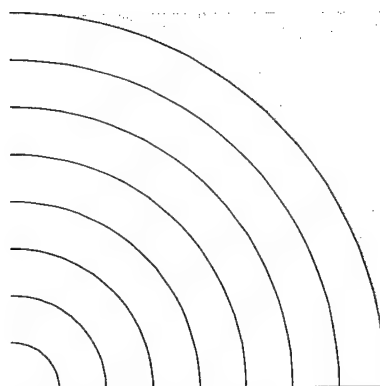


FIGURE 6.16. Results from VQ-based pass-band only restoration using separable DWT. (a) is original image, (b) is blurred image, $\text{SNR}=21.50$ dB, (c) is restored image, $\text{ISNR}=3.28$ dB, (d) is spectral correlation image.



(a)



(b)

FIGURE 6.17. Graphical representation of 3 level transform for (a) separable DWT and (b) RS-DWT.

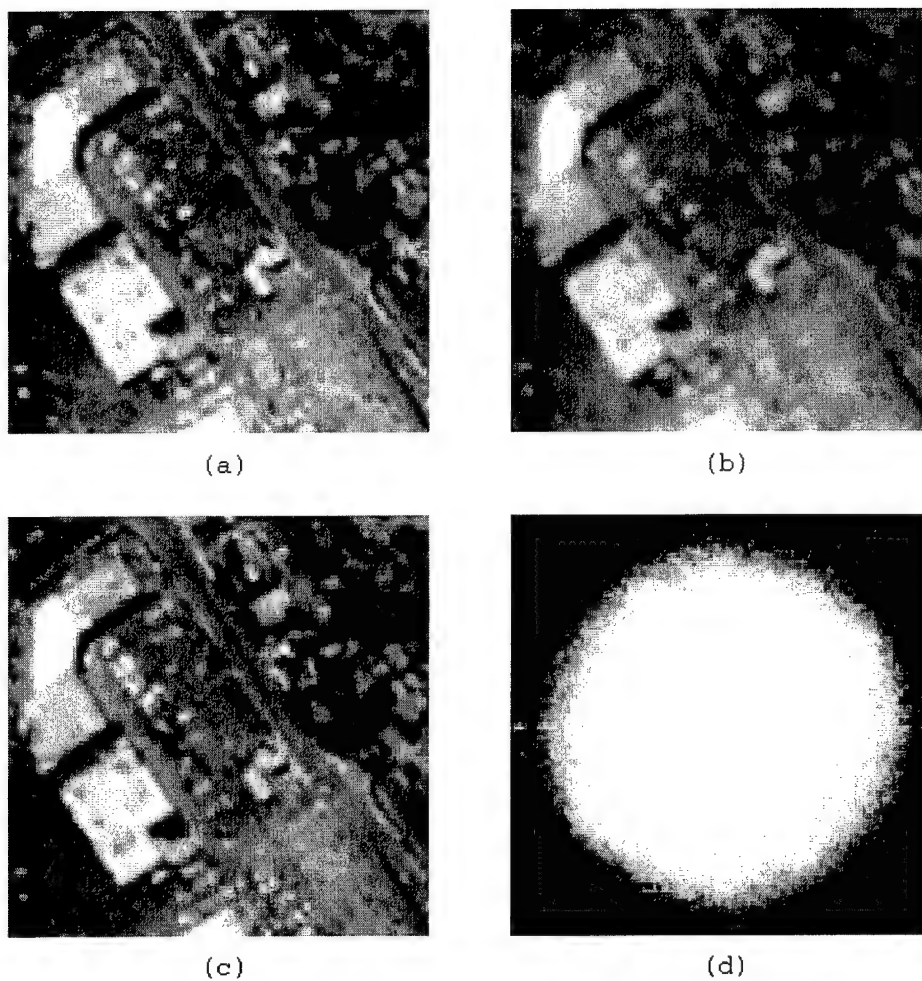


FIGURE 6.18. Results from VQ-based pass-band only restoration using RS-DWT. (a) is original image, (b) is blurred image, $\text{SNR}=21.62$ dB, (c) is restored image, $\text{ISNR}=6.75$ dB, (d) is spectral correlation image.

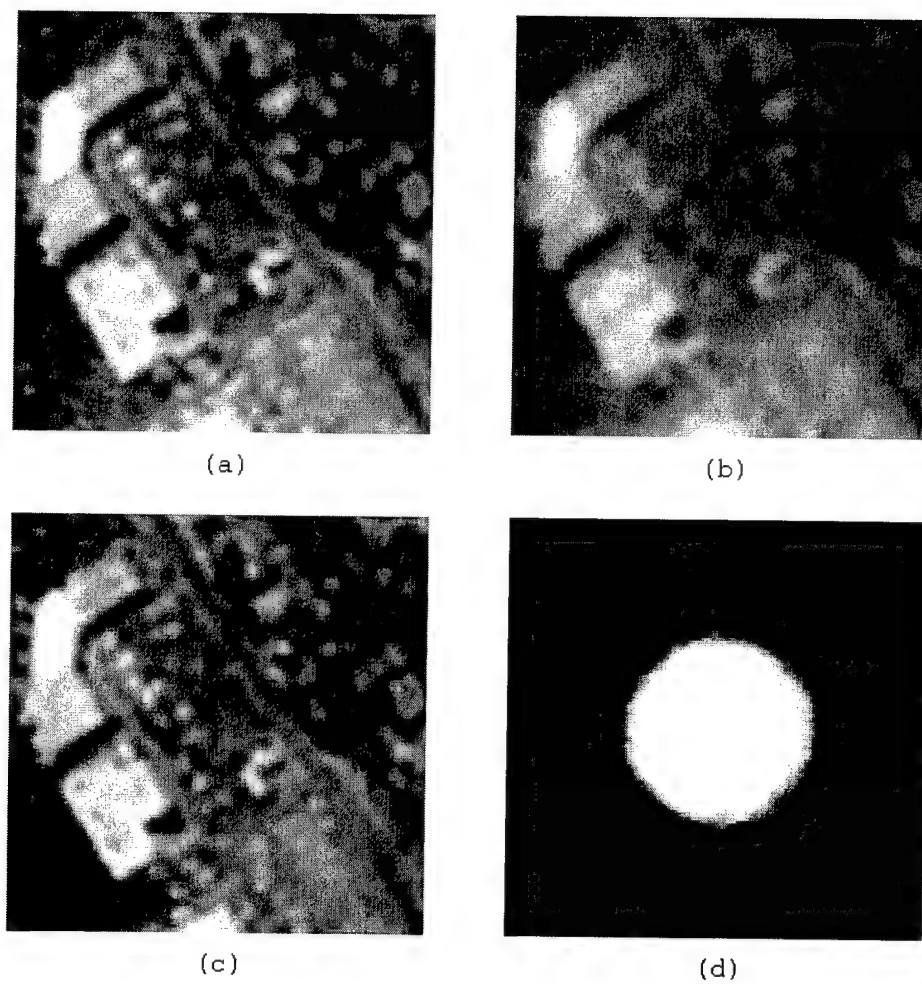


FIGURE 6.19. Results for super-resolution using separable DWT. (a) is original, (b) is blurred, SNR=17.11 dB, (c) is restored, ISNR=3.28 dB, (d) is spectral correlation plot.

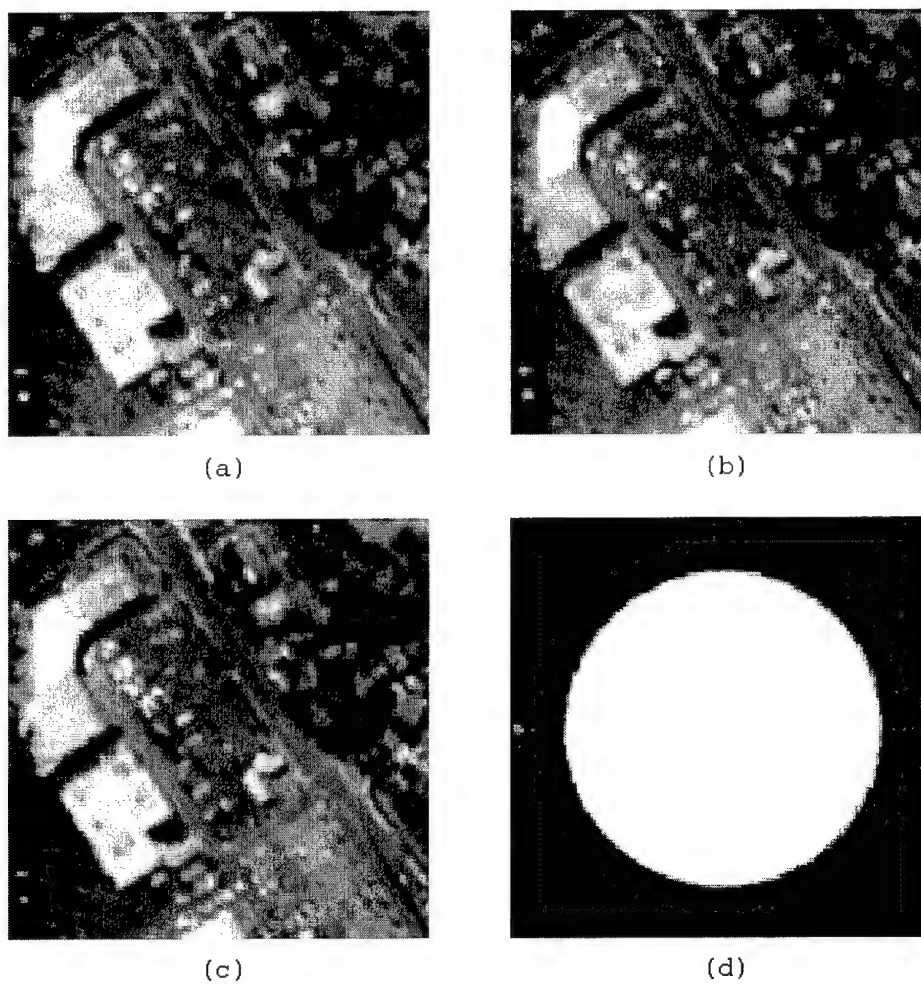


FIGURE 6.20. Results for super-resolution using RS-DWT uniform packet basis. (a) is original image, (b) is band-limited image $\text{SNR}=29.67$, (c) is restoration $\text{ISNR}=-0.81$ dB, and (d) is spectral correlation of restoration with original. The SR-ISNR is 0.15 dB.

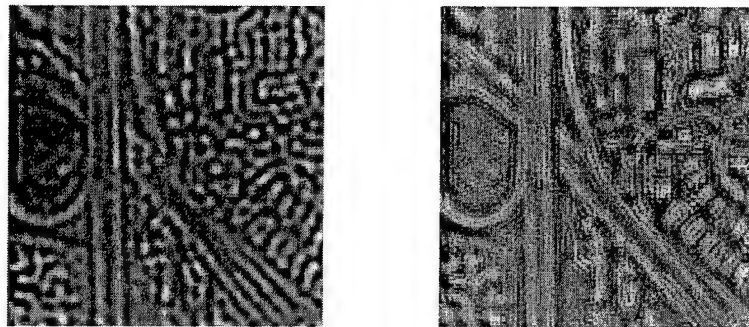


FIGURE 6.21. Images of wavelet subbands for the NLIVQ estimation (left) and truth data (right) of the same subband which is beyond the band-limit (super-resolution).

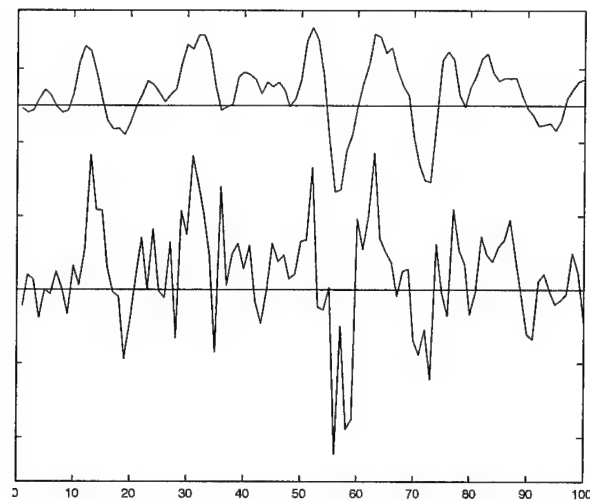


FIGURE 6.22. 1-D plot of the wavelet coefficients. Top is the NLIVQ based estimation, bottom is the truth data.

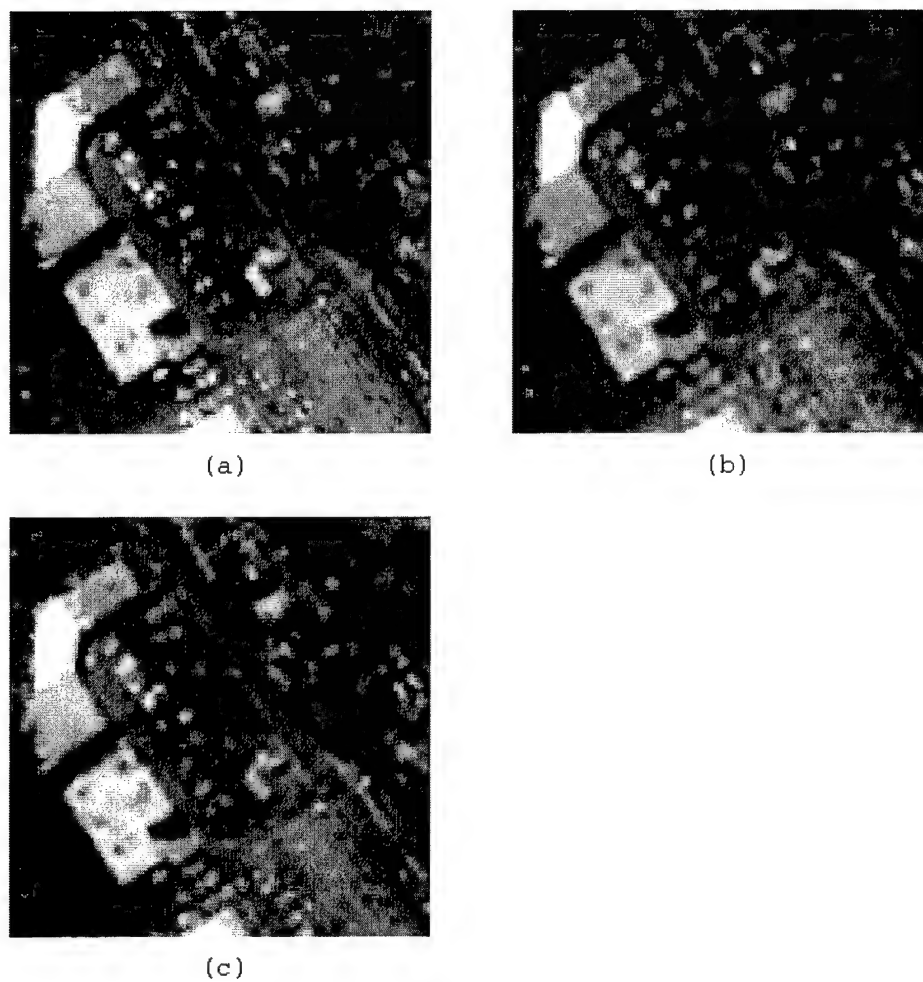


FIGURE 6.23. Results for DCT based NLIVQ estimation. (a) is the original, (b) is band-pass image $\text{SNR}=23.76$, and (c) is the NLIVQ restoration, $\text{SNR}=24.41$. The SR-ISNR is 0.95 dB.

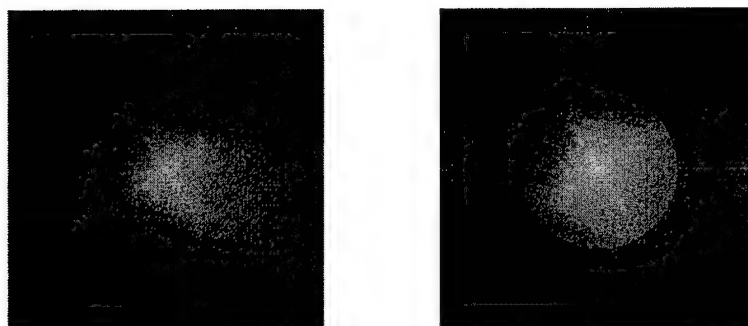


FIGURE 6.24. Log-compressed image of frequency content of true image (left) and NLIVQ estimate (right). Black circles show the band-pass cut-off.

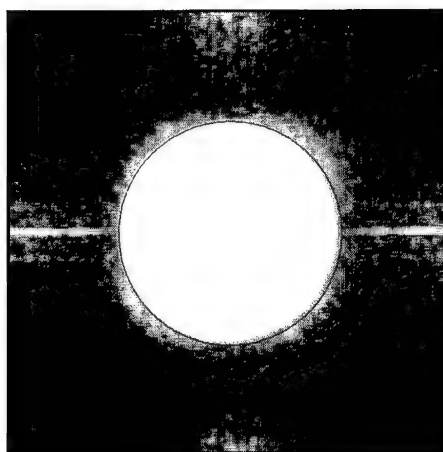


FIGURE 6.25. Complex correlation plot of NLIVQ estimation.

CHAPTER 7

CONCLUDING REMARKS

The goal of this research was to analyze image restoration and super-resolution from a multi-resolution perspective. In chapter 5, a new transform was introduced which was designed to solve the problem facing other wavelet domain image restoration techniques: how to account for the non-separable and circular nature of common imaging systems. The transform was motivated and presented as a way to maintain radial symmetry in the frequency domain while still taking advantage of the superior denoising performance of wavelets. When applied to blurred and noisy imagery, the results were better than Wiener Filtering (which inherently adapts to whatever symmetry occurs in frequency space) in terms of ISNR. From a subjective perspective, the RS-DWT restorations also tended to remove noise better in smooth areas of the imagery while maintaining the ability to model sharp transitions. One of the primary reasons for this is the spatial locality of wavelets versus the global nature of Fourier components. When Wiener filtering of the wavelet coefficients was also used, the results, in terms of ISNR, were close to the optimal LSI solution, the Wiener filter given perfect knowledge of the noise and signal PSD's. Again, while a subjective call, visual comparison and 1-D plots demonstrate that the results from the RS-DWT (with no a priori knowledge other than blurring function) were superior.

It is expected that these results could be further improved by use of more detailed denoising approaches which take into account more of the relationship between coefficients such as a neighborhood weighting or Hidden Markov Models. The soft thresholding and Wiener filter denoising approaches are among the simplest that have been developed. Based on results in the literature that demonstrate superior performance from more complicated modeling of the denoising process, additional improvement

should be possible.

Super-resolution, defined as the meaningful restoration of the object spectrum beyond the optical system cut-off, was also investigated from a multi-resolution perspective. Super-resolution from a multi-resolution perspective is the estimation of a new subband containing the higher frequency object information. Simple experiments were conducted to show that only a very few coefficients can yield significant super-resolution, even without perfect knowledge of the magnitude. However, determination of location is the problem for this approach. Although in general, large wavelet coefficients persist through the scales, this is not consistent enough to locate their positions, especially in the case of textured imagery. To compensate for the issues associated with a direct measurement of the coefficient location, an indirect method of using the relationship of coefficients across scales to predict the unknown subband via NLIVQ was used. The results showed decent improvement in passband restoration, although still below what is possible by other algorithms such as the RS-DWT technique developed in chapter 5. In order to determine the feasibility of super-resolution and separate out the band-pass improvements, perfect band-pass knowledge was assumed. The ability to super-resolve beyond the known frequencies was not significant. At best, only modest improvements of the image were produced. While not a comprehensive study, these results point out the difficulty in super-resolving in the wavelet domain and mirrors some of the issues that have arisen in the field of image compression.

REFERENCES

- [1] J. W. Goodman, *Introduction to Fourier Optics*. New York: McGraw Hill, 2nd ed., 1996.
- [2] J. D. Gaskill, *Linear Systems, Fourier Transforms, and Optics*. New York: Wiley, 1978.
- [3] H. C. Andrews and B. R. Hunt, *Digital Image Restoration*. Englewood Cliffs, NJ: Prentice-Hall, 1976.
- [4] A. Jain, *Fundamentals of Digital Image Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1989.
- [5] A. Katsaggelos, ed., *Digital Image Restoration*. New York: Springer-Verlag, 1991.
- [6] H. H. Barrett and K. J. Meyers, *Image Science: Mathematical and Statistical Foundations*. Wiley, 2001.
- [7] M. Banham and A. Katsaggelos, "Digital image restoration," *IEEE Signal Processing Magazine*, vol. 14, pp. 24–41, March 1997.
- [8] E. Hecht, *Optics*. Addison-Wesley, 2nd ed., 1987.
- [9] A. Hillery and R. Chin, "Iterative Wiener filters for image restoration," *IEEE Trans. Signal Processing*, vol. 39, pp. 1892–1899, Aug 1991.
- [10] N. Nguyen and P. Milanfar, "A computationally efficient superresolution image reconstruction algorithm," *IEEE Trans. Image Processing*, vol. 10, pp. 573–583, April 2001.
- [11] J. Green and B. Hunt, "Improved restoration of space object imagery," *J. Opt. Soc. Am. A*, vol. 16, pp. 2859–2865, December 1999.
- [12] D. Sheppard, A. Bilgin, M. Nadar, B. Hunt, and M. Marcellin, "A vector quantizer for image restoration," *IEEE Trans. Image Processing*, vol. 7, pp. 119–124, January 1998.
- [13] G. Toraldo di Francia, "Resolving power and information," *J. Opt. Soc. Am.*, vol. 45, pp. 497–501, July 1955.
- [14] R. Gerchberg, "Super-resolution through error energy reduction," *Optica Acta*, vol. 21, pp. 709–720, September 1974.

- [15] A. Papoulis, "A new algorithm in spectral analysis and band-limited extrapolation," *IEEE Trans. on Circuits and Systems*, vol. CAS-22, pp. 735-742, September 1975.
- [16] W. H. Richardson, "Bayesian-based iterative method of image restoration," *J. Opt. Soc. Am.*, vol. 62, pp. 55-59, January 1972.
- [17] L. B. Lucy, "An iterative technique for the rectification of observed distributions," *Astron. J.*, vol. 79, pp. 745-754, June 1974.
- [18] B. R. Frieden, "Restoring with maximum likelihood and maximum entropy," *J. Opt. Soc. Am.*, vol. 62, pp. 511-518, April 1972.
- [19] B. R. Frieden, "Restoring with maximum entropy, II: Superresolution of photographs of diffraction-blurred impulses," *J. Opt. Soc. Am.*, vol. 62, pp. 1202-1210, October 1972.
- [20] S. F. Gull and J. Skilling, "Maximum entropy method in image processing," *Proc. IEEE*, vol. 131-F, pp. 646-659, October 1984.
- [21] B. R. Hunt and P. Sementilli, "Description of a Poisson imagery super-resolution algorithm," in *Astronomical Data Analysis software and Systems, I* (Worrall et al., ed.), vol. 25, San Francisco: Astronomical Society of the Pacific, 1992.
- [22] P. J. Sementilli, *Suppression of artifacts in super-resolved images*. PhD thesis, University of Arizona, 1993.
- [23] B. R. Hunt, "Super-resolution of images: algorithms, principles, performance," *Int. J. Imaging Sys. and Tech.*, vol. 6, pp. 297-304, Winter 1995.
- [24] C. Miller, "Reconstruction and super resolution of dilute aperture imagery," Master's thesis, University of Arizona, Tucson, AZ, 1996.
- [25] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*. NY: McGraw-Hill, 3rd ed., 1991.
- [26] D. Hubel and T. Wiesel, "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex," *J. of Physiology*, vol. 160, 1962.
- [27] J. Daugmann, "Two-dimensional spectral analysis of cortical receptive field profile," *Vision Research*, vol. 20, pp. 847-856, 1980.
- [28] T. S. Lee, "Image representation using 2D Gabor wavelets," *IEEE Trans. PAMI*, vol. 18, pp. 959-971, Oct 1996.

- [29] P. Burt and E. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Trans. Comm.*, vol. 31, pp. 532–540, April 1983.
- [30] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. on PAMI*, vol. 11, pp. 674–693, July 1989.
- [31] S. G. Mallat, *A Wavelet Tour of Signal Processing*. Academic Press, 2nd ed., 1999.
- [32] C. Burrus, R. Gopinath, and H. Guo, *Introduction to Wavelets and Wavelet Transforms: A Primer*. Prentice Hall, 1998.
- [33] R. Gopinath and C. Burrus, "On the moments of the scaling function ψ_0 ," in *Proc of ISCAS'92, San Diego, CA*, pp. 963–966, May 1992.
- [34] M. J. Shensa, "The discrete wavelet transform: Wedding the A Trouns and Mallat algorithms," *IEEE Trans. Signal Processing*, vol. 40, pp. 2464–2482, October 1992.
- [35] D. Donoho and I. Johnstone, "Ideal spatial adaptation via wavelet shrinkage," *Biometrika*, vol. 81, pp. 425–455, December 1994.
- [36] R. Coifman and D. Donoho in *Wavelets and Statistics* (A. Antoniadis and G. Oppenheim, eds.), ch. Translation Invariant De-noising, Springer-Verlag, 1995.
- [37] S. G. Chang, B. Yu, and M. Vetterli, "Spatially adaptive wavelet thresholding with context modeling for image denoising," in *Proc. of ICIP'98, Chicago, IL*, vol. I, pp. 535–539, October 1998.
- [38] S. G. Chang, B. Yu, and M. Vetterli, "Spatially adaptive wavelet thresholding with context modeling for image denoising," *IEEE Trans. Image Processing*, vol. 9, pp. 1522–1531, September 2000.
- [39] S. P. Ghael, A. M. Sayeed, and R. G. Baraniuk, "Improved wavelet denoising via empirical Wiener filtering," in *Proc SPIE, San Diego, CA*, vol. 3169, pp. 389–399, July 1997.
- [40] H. Choi and R. Baraniuk, "Analysis of wavelet domain Wiener filters," in *Proc IEEE Time-Frequency and Time-Scale Analysis Conference*, pp. 613–616, 1998.
- [41] H. Choi and R. Baraniak, "Wavelet statistical models and Besov spaces," in *Proc of SPIE, Denver CO*, vol. 3813, pp. 489–501, July 1999.
- [42] J. Liu and P. Moulin, "Complexity regularized image restoration," in *Proc. ICIP'97, Santa Barabara, CA*, pp. 555–559, October 1997.

- [43] P. Moulin and J. Liu, "Analysis of multiresolution image denoising schemes using generalized-Gaussian and complexity priors," *IEEE Trans. on Info. Theory*, vol. 45, pp. 909–919, April 1999.
- [44] M. S. Crouse, R. D. Nowak, and R. G. Baraniuk, "Wavelet-based statistical signal processing using hidden Markov models," *IEEE Trans. Signal Processing*, vol. 46, pp. 886–902, Apr. 1998.
- [45] J. Liu and P. Moulin, "Image denoising based on scale-space mixture modeling of wavelet coefficients," in *Proc ICIP'99, Kobe, Japan*, vol. 1, pp. 386–390, 1999.
- [46] J. Lui and P. Moulin, "Information theoretic analysis of interscale and intrascale dependencies between image wavelet coefficients," *IEEE Trans. Image Processing*, vol. 10, pp. 1647–1658, November 2001.
- [47] R. Nowak and R. Baraniuk, "Wavelet-domain filtering for photon imaging systems," in *Proc. SPIE, San Diego, CA*, vol. 3169, pp. 55–66, 1997.
- [48] R. Nowak, "Multiscale hidden Markov models for photon limited imaging," in *Proc SPIE, Denver, CO*, vol. 3816, pp. 321–332, 1999.
- [49] K. Timmerman and R. Nowak, "Multi-scale modeling and estimation of Poisson processes with application to photon-limited imaging," *IEEE Trans. Info. Theory*, vol. 45, pp. 846–862, April 1999.
- [50] M. K. Mihcak, I. Kozintsev, and K. Ramchandran, "Spatially adaptive statistical modeling of wavelet image coefficients and its application to denoising," in *Proc. ICASSP'99*, vol. 6, pp. 3253–3256, 1999.
- [51] J. Kalifa, S. Mallat, and B. Rouge, "Image deconvolution in mirror wavelet bases," in *Proc. of ICIP'98, Chicago, IL*, vol. 1, pp. 565–569, 1998.
- [52] J. Kalifa, S. Mallat, and B. Rouge, "Minimax deconvolution in mirror wavelet bases," *submitted to IEEE Trans. Image Processing*.
- [53] R. Neelamani, H. Choi, and R. Baraniuk, "Wavelet-based deconvolution for ill-conditioned systems," *submitted to IEEE Trans. Image Processing*.
- [54] M. Banham and A. Katsaggelos, "Spatially adaptive wavelet-based multiscale image restoration," *IEEE Trans. Image Processing*, vol. 5, pp. 619–634, April 1996.
- [55] Y. Wan and R. Nowak, "A Bayesian multiscale approach to joint image restoration and edge detection," in *Proc SPIE, Denver, CO*, vol. 3813, pp. 73–84, July 1999.

- [56] A. Jalobeanu, L. Blanc-Feraud, and J. Zerubia, "Satellite deconvolution using complex wavelet packets," in *Proc ICIP'00, Vancouver, Canada*, vol. 3, pp. 809–812, 2000.
- [57] S. G. Chang, Z. Cvetkovich, and M. Vetterli, "Resolution enhancement of images using wavelet transform extrema extrapolation," in *Proc. ICASSP'95*, pp. 2379–2382, 1995.
- [58] Z. Cvetkovic and M. Vetterli, "Discrete-time wavelet extrema representation: Design and consistent reconstruction," *IEEE Trans. Signal Processing*, vol. 43, pp. 681–693, March 1995.
- [59] W. K. Carey, D. Chuang, and S. Hemami, "Regularity-preserving image interpolation," in *Proc ICASSP'97*, pp. 901–903, 1997.
- [60] W. K. Carey, D. Chuang, and S. Hemami, "Regularity-preserving image interpolation," *IEEE Trans. Image Processing*, vol. 8, pp. 1293–1297, Sep. 1999.
- [61] Y. Itoh, Y. Izumi, and Y. Tanaka, "Image enhancement based on estimation of high resolution component using wavelet transform," in *Proc ICIP'99, Kobe, Japan*, pp. 489–493, 1999.
- [62] J. Nunez, X. Otazu, O. Fore, A. Prades, V. Pala, and R. Arbiol, "Image fusion with additive multiresolution wavelet decomposition. Applications to SPOT+Landsat images," *J. Opt. Soc. Am. A*, vol. 12, pp. 467–474, Mar. 1999.
- [63] P. Scheunders, "Multiscale edge representation applied to image fusion," in *Proc SPIE, San Diego, CA*, vol. 4119, August 2000.
- [64] D. Wei and S. Guo, "A new approach to the design of multidimensional nonseparable two-channel orthonormal filterbanks and wavelets," *IEEE Signal Proc. Letters*, vol. 7, pp. 327–330, November 2000.
- [65] J. Kovacevic and M. Vetterli, "Nonseparable two- and three-dimensional wavelets," *IEEE Trans. Signal Processing*, vol. 43, pp. 1269–1273, May 1995.
- [66] H. Choi and R. Baraniak, "Multiple basis wavelet denoising using Besov projections," in *Proc ICIP'99, Kobe, Japan*, vol. 1, pp. 595–599, 1999.
- [67] J. Villasenor, B. Belzer, and J. Liao, "Wavelet filter evaluation for image compression," *IEEE Trans. Image Processing*, vol. 4, pp. 1053–1060, August 1995.
- [68] A. Cohen, I. Daubechies, and J. Feauveau, "Biorthogonal bases of compactly supported wavelets," *Commun. Pure Appl. Math*, vol. 45, pp. 485–560, June 1992.

- [69] R. Neelamani, "Ward software, version 1.0, available at <http://www.dsp.rice.edu/software/ward.shtml>."
- [70] E. Candes and D. Donoho, "Curvelets, multiresolution representation, and scaling laws," in *Proc SPIE, (San Diego, CA)*, vol. 4119, 2000.
- [71] M. Do and M. Vetterli, "Pyramidal directional filter banks and curvelets," in *Proc ICIP'01 (Thessaloniki, Greece)*, 2001.
- [72] R. Bamberger and M. Smith, "A filter bank for the directional decomposition of images: Theory and design," *IEEE Trans. Signal Processing*, vol. 40, pp. 882–893, April 1992.
- [73] A. Deever and S. S. Hemami, "What's your sign?: Efficient sign coding for embedded wavelet image coding," in *Proc IEEE Data Compression Conf. (Snowbird, UT)*, pp. 273–282, 2000.
- [74] A. Gersho and R. Gray, *Vector Quantization and Signal Compression*. Norwell, MA: Kluwer Academic, 1992.
- [75] K. Panchapakesan, A. Bilgin, M. Marcellin, and B. Hunt, "Joint compression and restoration of images using wavelets and non-linear interpolative vector quantization," in *Proc IEEE ICASSP'98, Seattle, Washington*, pp. 2649–2652, 1998.
- [76] D. Sheppard, K. Panchapakesan, A. Bilgin, B. Hunt, and M. Marcellin, "Lapped nonlinear interpolative vector quantization and image super-resolution," *IEEE Trans. Image Processing*, vol. 9, pp. 295–298, February 2000.